

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | |
|---|--|---|
| (51) International Patent Classification ⁶ : C12N 15/90, 15/85, C12Q 1/68, C12N 5/10, 9/12, 15/13, C07K 16/28, C12N 15/12, C07K 14/705, G01N 33/53, C12N 15/62, C07K 19/00 | A1 | (11) International Publication Number: WO 98/41645 (43) International Publication Date: 24 September 1998 (24.09.98) |
| (21) International Application Number: PCT/US98/03935 (22) International Filing Date: 9 March 1998 (09.03.98) (30) Priority Data: 08/819,866 14 March 1997 (14.03.97) US 09/023,715 13 February 1998 (13.02.98) US (71) Applicant: IDEC PHARMACEUTICALS CORPORATION [US/US]; 11011 Torreyana Road, San Diego, CA 92121 (US). (72) Inventors: REFF, Mitchell, E.; 4166 Combe Way, San Diego, CA 92122 (US). BARNETT, Richard, Spence; 306 Belmont Court, San Marcos, CA 92069 (US). McLACHLAN, Karen, Retta; Apartment B6, 766 South Nardo, Solana Beach, CA 92075 (US). (74) Agents: GESS, E., Joseph et al.; Burns, Doane, Swecker & Mathis L.L.P., P.O. Box 1404, Alexandria, VA 22313-1404 (US). | (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i> | |
| (54) Title: METHOD FOR INTEGRATING GENES AT SPECIFIC SITES IN MAMMALIAN CELLS VIA HOMOLOGOUS RECOMBINATION AND VECTORS FOR ACCOMPLISHING THE SAME (57) Abstract A method for achieving site specific integration of a desired DNA at a target site in a mammalian cell via homologous recombination is described. This method provides for the reproducible selection of cell lines wherein a desired DNA is integrated at a predetermined transcriptionally active site previously marked with a marker plasmid. The method is particularly suitable for the production of mammalian cell lines which secrete mammalian proteins at high levels, in particular immunoglobulins. Vectors and vector combinations for use in the subject cloning method are also provided. | | |

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|----|--------------------------|----|--|----|--|----|--------------------------|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | ML | Mali | TR | Turkey |
| BG | Bulgaria | HU | Hungary | MN | Mongolia | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MR | Mauritania | UA | Ukraine |
| BR | Brazil | IL | Israel | MW | Malawi | UG | Uganda |
| BY | Belarus | IS | Iceland | MX | Mexico | US | United States of America |
| CA | Canada | IT | Italy | NE | Niger | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NL | Netherlands | VN | Viet Nam |
| CG | Congo | KE | Kenya | NO | Norway | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NZ | New Zealand | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | PL | Poland | | |
| CM | Cameroon | KR | Republic of Korea | PT | Portugal | | |
| CN | China | KZ | Kazakhstan | RO | Romania | | |
| CU | Cuba | LC | Saint Lucia | RU | Russian Federation | | |
| CZ | Czech Republic | LI | Liechtenstein | SD | Sudan | | |
| DE | Germany | LK | Sri Lanka | SE | Sweden | | |
| DK | Denmark | LR | Liberia | SG | Singapore | | |
| EE | Estonia | | | | | | |

Title of the Invention

METHOD FOR INTEGRATING GENES AT SPECIFIC SITES IN MAMMALIAN CELLS VIA
HOMOLOGOUS RECOMBINATION AND VECTORS FOR ACCOMPLISHING THE SAME

5

Field of the Invention

The present invention relates to a process of targeting the integration of a desired exogenous DNA to a specific location within the genome of a mammalian cell.

10 More specifically, the invention describes a novel method for identifying a transcriptionally active target site ("hot spot") in the mammalian genome, and inserting a desired DNA at this site via homologous recombination. The invention also optionally provides the ability for

15 gene amplification of the desired DNA at this location by co-integrating an amplifiable selectable marker, e.g., DHFR, in combination with the exogenous DNA. The invention additionally describes the construction of novel vectors suitable for accomplishing the above, and

20 further provides mammalian cell lines produced by such methods which contain a desired exogenous DNA integrated at a target hot spot.

- 2 -

Background

Technology for expressing recombinant proteins in both prokaryotic and eukaryotic organisms is well established. Mammalian cells offer significant advantages over bacteria or yeast for protein production, resulting from their ability to correctly assemble, glycosylate and post-translationally modify recombinantly expressed proteins. After transfection into the host cells, recombinant expression constructs can be maintained as extrachromosomal elements, or may be integrated into the host cell genome. Generation of stably transfected mammalian cell lines usually involves the latter; a DNA construct encoding a gene of interest along with a drug resistance gene (dominant selectable marker) is introduced into the host cell, and subsequent growth in the presence of the drug allows for the selection of cells that have successfully integrated the exogenous DNA. In many instances, the gene of interest is linked to a drug resistant selectable marker which can later be subjected to gene amplification. The gene encoding dihydrofolate reductase (DHFR) is most commonly used for this purpose. Growth of cells in the presence of methotrexate, a competitive inhibitor of DHFR, leads to increased DHFR production by means of amplification of the DHFR gene. As flanking regions of DNA will also become amplified, the resultant coamplification of a DHFR linked gene in the transfected cell line can lead to increased protein

- 3 -

production, thereby resulting in high level expression of the gene of interest.

While this approach has proven successful, there are a number of problems with the system because of the random nature of the integration event. These problems exist because expression levels are greatly influenced by the effects of the local genetic environment at the gene locus, a phenomena well documented in the literature and generally referred to as "position effects" (for example, see Al-Shawi et al, *Mol. Cell. Biol.*, 10:1192-1198 (1990); Yoshimura et al, *Mol. Cell. Biol.*, 7:1296-1299 (1987)). As the vast majority of mammalian DNA is in a transcriptionally inactive state, random integration methods offer no control over the transcriptional fate of the integrated DNA. Consequently, wide variations in the expression level of integrated genes can occur, depending on the site of integration. For example, integration of exogenous DNA into inactive, or transcriptionally "silent" regions of the genome will result in little or no expression. By contrast integration into a transcriptionally active site may result in high expression.

Therefore, when the goal of the work is to obtain a high level of gene expression, as is typically the desired outcome of genetic engineering methods, it is generally necessary to screen large numbers of transfectants to find such a high producing clone.

- 4 -

Additionally, random integration of exogenous DNA into the genome can in some instances disrupt important cellular genes, resulting in an altered phenotype. These factors can make the generation of high expressing
5 stable mammalian cell lines a complicated and laborious process.

Recently, our laboratory has described the use of DNA vectors containing translationally impaired dominant selectable markers in mammalian gene expression. (This
10 is disclosed in U.S. Serial No. 08/147,696 filed November 3, 1993, recently allowed).

These vectors contain a translationally impaired neomycin phosphotransferase (neo) gene as the dominant selectable marker, artificially engineered to contain an
15 intron into which a DHFR gene along with a gene or genes of interest is inserted. Use of these vectors as expression constructs has been found to significantly reduce the total number of drug resistant colonies produced, thereby facilitating the screening procedure in
20 relation to conventional mammalian expression vectors. Furthermore, a significant percentage of the clones obtained using this system are high expressing clones. These results are apparently attributable to the modifications made to the neo selectable marker. Due to
25 the translational impairment of the neo gene, transfected cells will not produce enough neo protein to survive drug selection, thereby decreasing the overall

- 5 -

number of drug resistant colonies. Additionally, a higher percentage of the surviving clones will contain the expression vector integrated into sites in the genome where basal transcription levels are high, resulting in overproduction of neo, thereby allowing the cells to overcome the impairment of the neo gene. Concomitantly, the genes of interest linked to neo will be subject to similar elevated levels of transcription. This same advantage is also true as a result of the artificial intron created within neo; survival is dependent on the synthesis of a functional neo gene, which is in turn dependent on correct and efficient splicing of the neo introns. Moreover, these criteria are more likely to be met if the vector DNA has integrated into a region which is already highly transcriptionally active.

Following integration of the vector into a transcriptionally active region, gene amplification is performed by selection for the DHFR gene. Using this system, it has been possible to obtain clones selected using low levels of methotrexate (50nM), containing few (<10) copies of the vector which secrete high levels of protein (>55pg/cell/day). Furthermore, this can be achieved in a relatively short period of time. However, the success in amplification is variable. Some transcriptionally active sites cannot be amplified and

- 6 -

therefore the frequency and extent of amplification from a particular site is not predictable.

Overall, the use of these translationally impaired vectors represents a significant improvement over other
5 methods of random integration. However, as discussed, the problem of lack of control over the integration site remains a significant concern.

One approach to overcome the problems of random integration is by means of gene targeting, whereby the
10 exogenous DNA is directed to a specific locus within the host genome. The exogenous DNA is inserted by means of homologous recombination occurring between sequences of DNA in the expression vector and the corresponding homologous sequence in the genome. However, while this
15 type of recombination occurs at a high frequency naturally in yeast and other fungal organisms, in higher eukaryotic organisms it is an extremely rare event. In mammalian cells, the frequency of homologous versus non-homologous (random integration) recombination is reported to range from 1/100 to 1/5000 (for example, see
20 Capecchi, *Science*, 244:1288-1292 (1989); Morrow and Kucherlapati, *Curr. Op. Biotech.*, 4:577-582 (1993)).

One of the earliest reports describing homologous recombination in mammalian cells comprised an artificial
25 system created in mouse fibroblasts (Thomas et al, *Cell*, 44:419-428 (1986)). A cell line containing a mutated, non-functional version of the neo gene integrated into

- 7 -

the host genome was created, and subsequently targeted with a second non-functional copy of neo containing a different mutation. Reconstruction of a functional neo gene could occur only by gene targeting. Homologous recombina-
5 nts were identified by selecting for G418 resistant cells, and confirmed by analysis of genomic DNA isolated from the resistant clones.

Recently, the use of homologous recombination to replace the heavy and light immunoglobulin genes at
10 endogenous loci in antibody secreting cells has been reported. (U.S. Patent No. 5,202,238, Fell et al, (1993).) However, this particular approach is not widely applicable, because it is limited to the production of immunoglobulins in cells which
15 endogenously express immunoglobulins, e.g., B cells and myeloma cells. Also, expression is limited to single copy gene levels because co-amplification after homologous recombination is not included. The method is further complicated by the fact that two separate
20 integration events are required to produce a functional immunoglobulin: one for the light chain gene followed by one for the heavy chain gene.

An additional example of this type of system has been reported in NS/O cells, where recombinant
25 immunoglobulins are expressed by homologous recombination into the immunoglobulin gamma 2A locus (Hollis et al, international patent application #

- 8 -

PCT/IB95 (00014).) Expression levels obtained from this site were extremely high - on the order of 20pg/cell/day from a single copy integrant. However, as in the above example, expression is limited to this level because an
5 amplifiable gene is not contegrated in this system. Also, other researchers have reported aberrant glycosylation of recombinant proteins expressed in NS/O cells (for example, see Flesher et al, *Biotech. and Bioeng.*, 48:399-407 (1995)), thereby limiting the
10 applicability of this approach.

The cre-loxP recombination system from bacteriophage P1 has recently been adapted and used as a means of gene targeting in eukaryotic cells. Specifically, the site specific integration of exogenous
15 DNA into the Chinese hamster ovary (CHO) cell genome using cre recombinase and a series of lox containing vectors have been described. (Fukushige and Sauer, *Proc. Natl. Acad. Sci. USA*, 89:7905-7909 (1992).) This system is attractive in that it provides for
20 reproducible expression at the same chromosomal location. However, no effort was made to identify a chromosomal site from which gene expression is optimal, and as in the above example, expression is limited to single copy levels in this system. Also, it is
25 complicated by the fact that one needs to provide for expression of a functional recombinase enzyme in the mammalian cell.

- 9 -

The use of homologous recombination between an introduced DNA sequence and its endogenous chromosomal locus has also been reported to provide a useful means of genetic manipulation in mammalian cells, as well as
5 in yeast cells. (See e.g., Bradley et al, *Meth. Enzymol.*, 223:855-879 (1993); Capecchi, *Science*, 244:1288-1292 (1989); Rothstein et al, *Meth. Enzymol.*, 194:281-301 (1991)). To date, most mammalian gene targeting studies have been directed toward gene
10 disruption ("knockout") or site-specific mutagenesis of selected target gene loci in mouse embryonic stem (ES) cells. The creation of these "knockout" mouse models has enabled scientists to examine specific structure-function issues and examine the biological
15 importance of a myriad of mouse genes. This field of research also has important implications in terms of potential gene therapy applications.

Also, vectors have recently been reported by Cell-tech (Kent, U.K.) which purportedly are targeted to
20 transcriptionally active sites in NSO cells, which do not require gene amplification (Peakman et al, *Hum. Antibod. Hybridomas*, 5:65-74 (1994)). However, levels of immunoglobulin secretion in these unamplified cells have not been reported to exceed 20pg/cell/day, while in
25 amplified CHO cells, levels as high as 100pg/cell/day can be obtained (Id.).

- 10 -

It would be highly desirable to develop a gene targeting system which reproducibly provided for the integration of exogenous DNA into a predetermined site in the genome known to be transcriptionally active.

5 Also, it would be desirable if such a gene targeting system would further facilitate co-amplification of the inserted DNA after integration. The design of such a system would allow for the reproducible and high level expression of any cloned gene of interest in a mammalian
10 cell, and undoubtedly would be of significant interest to many researchers.

In this application, we provide a novel mammalian expression system, based on homologous recombination occurring between two artificial substrates contained in
15 two different vectors. Specifically, this system uses a combination of two novel mammalian expression vectors, referred to as a "marking" vector and a "targeting" vector.

Essentially, the marking vector enables the identification and marking of a site in the mammalian genome
20 which is transcriptionally active, i.e., a site at which gene expression levels are high. This site can be regarded as a "hot spot" in the genome. After integration of the marking vector, the subject expression system enables another DNA to be integrated at this site,
25 i.e., the targeting vector, by means of homologous recombination occurring between DNA sequences common to

- 11 -

both vectors. This system affords significant advantages over other homologous recombination systems.

Unlike most other homologous systems employed in mammalian cells, this system exhibits no background. Therefore, cells which have only undergone random integration of the vector do not survive the selection. Thus, any gene of interest cloned into the targeting plasmid is expressed at high levels from the marked hot spot. Accordingly, the subject method of gene expression substantially or completely eliminates the problems inherent to systems of random integration, discussed in detail above. Moreover, this system provides reproducible and high level expression of any recombinant protein at the same transcriptionally active site in the mammalian genome. In addition, gene amplification may be effected at this particular transcriptionally active site by including an amplifiable dominant selectable marker (e.g. DHFR) as part of the marking vector.

Objects of the Invention

Thus, it is an object of the invention to provide an improved method for targeting a desired DNA to a specific site in a mammalian cell.

It is a more specific object of the invention to provide a novel method for targeting a desired DNA to a specific site in a mammalian cell via homologous recombination.

- 12 -

It is another specific object of the invention to provide novel vectors for achieving site specific integration of a desired DNA in a mammalian cell.

It is still another object of the invention to
5 provide novel mammalian cell lines which contain a desired DNA integrated at a predetermined site which provides for high expression.

It is a more specific object of the invention to provide a novel method for achieving site specific integration of a desired DNA in a Chinese hamster ovary
10 (CHO) cell.

It is another more specific object of the invention to provide a novel method for integrating immunoglobulin genes, or any other genes, in mammalian cells at
15 predetermined chromosomal sites that provide for high expression.

It is another specific object of the invention to provide novel vectors and vector combinations suitable for integrating immunoglobulin genes into mammalian
20 cells at predetermined sites that provide for high expression.

It is another object of the invention to provide mammalian cell lines which contain immunoglobulin genes integrated at predetermined sites that provide for high
25 expression.

It is an even more specific object of the invention to provide a novel method for integrating immunoglobulin

- 13 -

genes into CHO cells that provide for high expression, as well as novel vectors and vector combinations that provide for such integration of immunoglobulin genes into CHO cells.

5 In addition, it is a specific object of the invention to provide novel CHO cell lines which contain immunoglobulin genes integrated at predetermined sites that provide for high expression, and have been amplified by methotrexate selection to secrete even greater amounts
10 of functional immunoglobulins.

Brief Description of the Figures

Figure 1 depicts a map of a marking plasmid according to the invention referred to as Desmond. The plasmid is shown in circular form (1a) as well as a
15 linearized version used for transfection (1b).

Figure 2(a) shows a map of a targeting plasmid referred to "Molly". Molly is shown here encoding the anti-CD20 immunoglobulin genes, expression of which is described in Example 1.

20 Figure 2(b) shows a linearized version of Molly, after digestion with the restriction enzymes KpnI and PacI. This linearized form was used for transfection.

Figure 3 depicts the potential alignment between Desmond sequences integrated into the CHO genome, and
25 incoming targeting Molly sequences. One potential ar-

- 14 -

rangement of Molly integrated into Desmond after homologous recombination is also presented.

Figure 4 shows a Southern analysis of single copy Desmond clones. Samples are as follows:

- 5 Lane 1: λ HindIII DNA size marker
- Lane 2: Desmond clone 10F3
- Lane 3: Desmond clone 10C12
- Lane 4: Desmond clone 15C9
- Lane 5: Desmond clone 14B5
- 10 Lane 6: Desmond clone 9B2

Figure 5 shows a Northern analysis of single copy Desmond clones. Samples are as follows: Panel A: northern probed with CAD and DHFR probes, as indicated on the figure. Panel B: duplicate northern, probed with CAD and HisD probes, as indicated. The RNA samples loaded in panels A and B are as follows:

- Lane 1: clone 9B2, lane 2; clone 10C12, lane 3; clone 14B5, lane 4; clone 15C9, lane 5; control RNA from CHO transfected with a HisD and DHFR containing plasmid,
- 15 lane 6; untransfected CHO.
- 20

Figure 6 shows a Southern analysis of clones resulting from the homologous integration of Molly into Desmond. Samples are as follows:

- Lane 1: λ HindIII DNA size markers, Lane 2: 20F4, lane 3;
- 25 5F9, lane 4; 21C7, lane 5; 24G2, lane 6; 25E1, lane 7;
- 28C9, lane 8; 29F9, lane 9; 39G11, lane 10; 42F9, lane 11; 50G10, lane 12; Molly plasmid DNA, linearized with

- 15 -

BglIII(top band) and cut with BglIII and KpnI (lower band), lane 13; untransfected Desmond.

Figures 7A through 7G contain the Sequence Listing for Desmond.

5 Figures 8A through 8I contain the Sequence Listing for Molly-containing anti-CD20.

Figure 9 contains a map of the targeting plasmid, "Mandy," shown here encoding anti-CD23 genes, the expression of which is disclosed in Example 5.

10 Figures 10A through 10N contain the sequence listing of "Mandy" containing the anti-CD23 genes as disclosed in Example 5.

Detailed Description of the Invention

15 The invention provides a novel method for integrating a desired exogenous DNA at a target site within the genome of a mammalian cell via homologous recombination. Also, the invention provides novel vectors for achieving the site specific integration of a DNA at a target site in the genome of a mammalian cell.

20 More specifically, the subject cloning method provides for site specific integration of a desired DNA in a mammalian cell by transfection of such cell with a "marker plasmid" which contains a unique sequence that is foreign to the mammalian cell genome and which
25 provides a substrate for homologous recombination, followed by transfection with a "target plasmid" containing

- 16 -

a sequence which provides for homologous recombination with the unique sequence contained in the marker plasmid, and further comprising a desired DNA that is to be integrated into the mammalian cell. Typically, the
5 integrated DNA will encode a protein of interest, such as an immunoglobulin or other secreted mammalian glycoprotein.

The exemplified homologous recombination system uses the neomycin phosphotransferase gene as a dominant
10 selectable marker. This particular marker was utilized based on the following previously published observations;

(i) the demonstrated ability to target and restore function to a mutated version of the neo gene (cited
15 earlier) and

(ii) our development of translationally impaired expression vectors, in which the neo gene has been artificially created as two exons with a gene of interest inserted in the intervening intron; neo exons are correctly spliced and translated in vivo, producing a functional protein and thereby conferring G418 resistance on the resultant cell population. In this application, the
20 neo gene is split into three exons. The third exon of neo is present on the "marker" plasmid and becomes integrated into the host cell genome upon integration of the
25 marker plasmid into the mammalian cells. Exons 1 and 2 are present on the targeting plasmid, and are separated

- 17 -

by an intervening intron into which at least one gene of interest is cloned. Homologous recombination of the targeting vector with the integrated marking vector results in correct splicing of all three exons of the neo gene and thereby expression of a functional neo protein (as determined by selection for G418 resistant colonies). Prior to designing the current expression system, we had experimentally tested the functionality of such a triply spliced neo construct in mammalian cells. The results of this control experiment indicated that all three neo exons were properly spliced and therefore suggested the feasibility of the subject invention.

However, while the present invention is exemplified using the neo gene, and more specifically a triple split neo gene, the general methodology should be efficacious with other dominant selectable markers.

As discussed in greater detail *infra*, the present invention affords numerous advantages to conventional gene expression methods, including both random integration and gene targeting methods. Specifically, the subject invention provides a method which reproducibly allows for site-specific integration of a desired DNA into a transcriptionally active domain of a mammalian cell. Moreover, because the subject method introduces an artificial region of "homology" which acts as a unique substrate for homologous recombination and the

- 18 -

insertion of a desired DNA, the efficacy of subject invention does not require that the cell endogenously contain or express a specific DNA. Thus, the method is generically applicable to all mammalian cells, and can
5 be used to express any type of recombinant protein.

The use of a triply spliced selectable marker, e.g., the exemplified triply spliced neo construct, guarantees that all G418 resistant colonies produced will arise from a homologous recombination event (random
10 integrants will not produce a functional neo gene and consequently will not survive G418 selection). Thus, the subject invention makes it easy to screen for the desired homologous event. Furthermore, the frequency of additional random integrations in a cell that has under-
15 gone a homologous recombination event appears to be low.

Based on the foregoing, it is apparent that a significant advantage of the invention is that it substantially reduces the number of colonies that need be screened to identify high producer clones, i.e., cell
20 lines containing a desired DNA which secrete the corresponding protein at high levels. On average, clones containing integrated desired DNA may be identified by screening about 5 to 20 colonies (compared to several thousand which must be screened when using standard
25 random integration techniques, or several hundred using the previously described intronic insertion vectors) Additionally, as the site of integration was preselected

- 19 -

and comprises a transcriptionally active domain, all exogenous DNA expressed at this site should produce comparable, i.e. high levels of the protein of interest.

Moreover, the subject invention is further advantageous in that it enables an amplifiable gene to be inserted on integration of the marking vector. Thus, when a desired gene is targeted to this site via homologous recombination, the subject invention allows for expression of the gene to be further enhanced by gene amplification. In this regard, it has been reported in from the literature that different genomic sites have different capacities for gene amplification (Meinkoth et al, *Mol. Cell Biol.*, 7:1415-1424 (1987)). Therefore, this technique is further advantageous as it allows for the placement of a desired gene of interest at a specific site that is both transcriptionally active and easily amplified. Therefore, this should significantly reduce the amount of time required to isolate such high producers.

Specifically, while conventional methods for the construction of high expressing mammalian cell lines can take 6 to 9 months, the present invention allows for such clones to be isolated on average after only about 3-6 months. This is due to the fact that conventionally isolated clones typically must be subjected to at least three rounds of drug resistant gene amplification in order to reach satisfactory levels of gene expression.

- 20 -

As the homologously produced clones are generated from a preselected site which is a high expression site, fewer rounds of amplification should be required before reaching a satisfactory level of production.

5 Still further, the subject invention enables the reproducible selection of high producer clones wherein the vector is integrated at low copy number, typically single copy. This is advantageous as it enhances the stability of the clones and avoids other potential adverse side-effects associated with high copy number. As
10 described supra, the subject homologous recombination system uses the combination of a "marker plasmid" and a "targeting plasmid" which are described in more detail below.

15 The "marker plasmid" which is used to mark and identify a transcriptionally hot spot will comprise at least the following sequences:

 (i) a region of DNA that is heterologous or unique to the genome of the mammalian cell, which functions as
20 a source of homology, allows for homologous recombination (with a DNA contained in a second target plasmid). More specifically, the unique region of DNA (i) will generally comprise a bacterial, viral, yeast synthetic, or other DNA which is not normally present in the
25 mammalian cell genome and which further does not comprise significant homology or sequence identity to DNA contained in the genome of the mammalian cell.

- 21 -

Essentially, this sequence should be sufficiently different to mammalian DNA that it will not significantly recombine with the host cell genome via homologous recombination. The size of such unique DNA will generally be at least about 2 to 10 kilobases in size, or higher, more preferably at least about 10kb, as several other investigators have noted an increased frequency of targeted recombination as the size of the homology region is increased (Capecchi, *Science*, 244:1288-1292 (1989)).

The upper size limit of the unique DNA which acts as a site for homologous recombination with a sequence in the second target vector is largely dictated by potential stability constraints (if DNA is too large it may not be easily integrated into a chromosome and the difficulties in working with very large DNAs.

(ii) a DNA including a fragment of a selectable marker DNA, typically an exon of a dominant selectable marker gene. The only essential feature of this DNA is that it not encode a functional selectable marker protein unless it is expressed in association with a sequence contained in the target plasmid. Typically, the target plasmid will comprise the remaining exons of the dominant selectable marker gene (those not comprised in "targeting" plasmid). Essentially, a functional selectable marker should only be produced if homologous recombination occurs (resulting in the association and

- 22 -

expression of this marker DNA (i) sequence together with the portion(s) of the selectable marker DNA fragment which is (are) contained in the target plasmid).

As noted, the current invention exemplifies the use of the neomycin phosphotransferase gene as the dominant selectable marker which is "split" in the two vectors. However, other selectable markers should also be suitable, e.g., the Salmonella histidinol dehydrogenase gene, hygromycin phosphotransferase gene, herpes simplex virus thymidine kinase gene, adenosine deaminase gene, glutamine synthetase gene and hypoxanthine-guanine phosphoribosyl transferase gene.

(iii) a DNA which encodes a functional selectable marker protein, which selectable marker is different from the selectable marker DNA (ii). This selectable marker provides for the successful selection of mammalian cells wherein the marker plasmid is successfully integrated into the cellular DNA. More preferably, it is desirable that the marker plasmid comprise two such dominant selectable marker DNAs, situated at opposite ends of the vector. This is advantageous as it enables integrants to be selected using different selection agents and further enables cells which contain the entire vector to be selected. Additionally, one marker can be an amplifiable marker to facilitate gene amplification as discussed previously. Any of the

- 23 -

dominant selectable marker listed in (ii) can be used as well as others generally known in the art.

Moreover, the marker plasmid may optionally further comprise a rare endonuclease restriction site. This is potentially desirable as this may facilitate cleavage. If present, such rare restriction site should be situated close to the middle of the unique region that acts as a substrate for homologous recombination. Preferably such sequence will be at least about 12 nucleotides. The introduction of a double stranded break by similar methodology has been reported to enhance the frequency of homologous recombination. (Choulika et al, *Mol. Cell. Biol.*, 15:1968-1973 (1995)). However, the presence of such sequence is not essential.

The "targeting plasmid" will comprise at least the following sequences:

(1) the same unique region of DNA contained in the marker plasmid or one having sufficient homology or sequence identity therewith that said DNA is capable of combining via homologous recombination with the unique region (i) in the marker plasmid. Suitable types of DNAs are described supra in the description of the unique region of DNA (1) in the marker plasmid.

(2) The remaining exons of the dominant selectable marker, one exon of which is included as (ii) in the marker plasmid listed above. The essential features of this DNA fragment is that it result in a functional

- 24 -

(selectable) marker protein only if the target plasmid integrates via homologous recombination (wherein such recombination results in the association of this DNA with the other fragment of the selectable marker DNA contained in the marker plasmid) and further that it allow for insertion of a desired exogenous DNA. Typically, this DNA will comprise the remaining exons of the selectable marker DNA which are separated by an intron. For example, this DNA may comprise the first two exons of the neo gene and the marker plasmid may comprise the third exon (back third of neo).

(3) The target plasmid will also comprise a desired DNA, e.g., one encoding a desired polypeptide, preferably inserted within the selectable marker DNA fragment contained in the plasmid. Typically, the DNA will be inserted in an intron which is comprised between the exons of the selectable marker DNA. This ensures that the desired DNA is also integrated if homologous recombination of the target plasmid and the marker plasmid occurs. This intron may be naturally occurring or it may be engineered into the dominant selectable marker DNA fragment.

This DNA will encode any desired protein, preferably one having pharmaceutical or other desirable properties. Most typically the DNA will encode a mammalian protein, and in the current examples provided, an immunoglobulin or an immunoadhesin. However the

- 25 -

invention is not in any way limited to the production of immunoglobulins.

As discussed previously, the subject cloning method is suitable for any mammalian cell as it does not require for efficacy that any specific mammalian sequence or sequences be present. In general, such mammalian cells will comprise those typically used for protein expression, e.g., CHO cells, myeloma cells, COS cells, BHK cells, Sp2/0 cells, NIH 3T3 and HeLa cells. In the examples which follow, CHO cells were utilized. The advantages thereof include the availability of suitable growth medium, their ability to grow efficiently and to high density in culture, and their ability to express mammalian proteins such as immunoglobulins in biologically active form.

Further, CHO cells were selected in large part because of previous usage of such cells by the inventors for the expression of immunoglobulins (using the translationally impaired dominant selectable marker containing vectors described previously). Thus, the present laboratory has considerable experience in using such cells for expression. However, based on the examples which follow, it is reasonable to expect similar results will be obtained with other mammalian cells.

In general, transformation or transfection of mammalian cells according to the subject invention will be effected according to conventional methods. So that the

- 26 -

invention may be better understood, the construction of exemplary vectors and their usage in producing integrants is described in the examples below.

EXAMPLE 1

5 Design and Preparation of Marker
 and Targeting Plasmid DNA Vectors

The marker plasmid herein referred to as "Desmond" was assembled from the following DNA elements:

10 (a) Murine dihydrofolate reductase gene (DHFR),
incorporated into a transcription cassette, comprising the mouse beta globin promoter 5" to the DHFR start site, and bovine growth hormone poly adenylation signal 3" to the stop codon. The DHFR transcriptional cassette was isolated from TCAE6, an expression vector created
15 previously in this laboratory (Newman et al, 1992, *Bio-technology*, 10:1455-1460).

 (b) E. coli β -galactosidase gene - commercially available, obtained from Promega as pSV-b-galactosidase control vector, catalog # E1081.

20 (c) Baculovirus DNA, commercially available, purchased from Clontech as pBAKPAK8, cat # 6145-1.

 (d) Cassette comprising promoter and enhancer elements from Cytomegalovirus and SV40 virus. The cassette was generated by PCR using a derivative of expression
25 vector TCAE8 (Reff et al, *Blood*, 83:435-445 (1994)). The enhancer cassette was inserted within the baculo-

- 27 -

virus sequence, which was first modified by the insertion of a multiple cloning site.

(e) E. coli GUS (glucuronidase) gene, commercially available, purchased from Clontech as pB101, cat. # 6017-1.

(f) Firefly luciferase gene, commercially available, obtained from Promega as pGEM-Luc (catalog # E1541).

(g) S. typhimurium histidinol dehydrogenase gene (HisD). This gene was originally a gift from (Donahue et al, Gene, 18:47-59 (1982)), and has subsequently been incorporated into a transcription cassette comprising the mouse beta globin major promoter 5' to the gene, and the SV40 polyadenylation signal 3' to the gene.

The DNA elements described in (a)-(g) were combined into a pBR derived plasmid backbone to produce a 7.7kb contiguous stretch of DNA referred to in the attached figures as "homology". Homology in this sense refers to sequences of DNA which are not part of the mammalian genome and are used to promote homologous recombination between transfected plasmids sharing the same homology DNA sequences.

(h) Neomycin phosphotransferase gene from TN5 (Davis and Smith, Ann. Rev. Micro., 32:469-518 (1978)).

The complete neo gene was subcloned into pBluescript SK- (Stratagene catalog # 212205) to facilitate genetic manipulation. A synthetic linker was then inserted into

- 28 -

a unique PstI site occurring across the codons for amino acid 51 and 52 of neo. This linker encoded the necessary DNA elements to create an artificial splice donor site, intervening intron and splice acceptor site within the neo gene, thus creating two separate exons, presently referred to as neo exon 1 and 2. Neo exon 1 encodes the first 51 amino acids of neo, while exon 2 encodes the remaining 203 amino acids plus the stop codon of the protein A NotI cloning site was also created within the intron.

Neo exon 2 was further subdivided to produce neo exons 2 and 3. This was achieved as follows: A set of PCR primers were designed to amplify a region of DNA encoding neo exon 1, intron and the first 111 2/3 amino acids of exon2. The 3' PCR primer resulted in the introduction of a new 5' splice site immediately after the second nucleotide of the codon for amino acid 111 in exon 2, therefore generating a new smaller exon 2. The DNA fragment now encoding the original exon 1, intron and new exon 2 was then subcloned and propagated in a pBR based vector. The remainder of the original exon 2 was used as a template for another round of PCR amplification, which generated "exon3". The 5' primer for this round of amplification introduced a new splice acceptor site at the 5' side of the newly created exon 3, i.e. before the final nucleotide of the codon for amino acid 111. The resultant 3 exons of neo encode the

- 29 -

following information: exon 1 - the first 51 amino acids of neo; exon 2 - the next 111 2/3 amino acids, and exon 3 the final 91 1/3 amino acids plus the translational stop codon of the neo gene.

5 Neo exon 3 was incorporated along with the above mentioned DNA elements into the marking plasmid "Desmond". Neo exons 1 and 2 were incorporated into the targeting plasmid "Molly". The NotI cloning site created within the intron between exons 1 and 2 was used in
10 subsequent cloning steps to insert genes of interest into the targeting plasmid.

A second targeting plasmid "Mandy" was also generated. This plasmid is almost identical to "Molly" (some restriction sites on the vector have been changed)
15 except that the original HisD and DHFR genes contained in "Molly" were inactivated. These changes were incorporated because the Desmond cell line was no longer being cultured in the presence of Histidinol, therefore it seemed unnecessary to include a second copy of the
20 HisD gene. Additionally, the DHFR gene was inactivated to ensure that only a single DHFR gene, namely the one present in the Desmond marked site, would be amplifiable in any resulting cell lines. "Mandy" was derived from "Molly" by the following modifications:

25 (i) A synthetic linker was inserted in the middle of the DHFR coding region. This linker created a stop codon and shifted the remainder of the DHFR coding

- 30 -

region out of frame, therefore rendering the gene nonfunctional.

(ii) A portion of the HisD gene was deleted and replaced with a PCR generated HisD fragment lacking the promoter and start codon of the gene.

Figure 1 depicts the arrangement of these DNA elements in the marker plasmid "Desmond". Figure 2 depicts the arrangement of these elements in the first targeting plasmid, "Molly". Figure 3 illustrates the possible arrangement in the CHO genome, of the various DNA elements after targeting and integration of Molly DNA into Desmond marked CHO cells. Figure 9 depicts the targeting plasmid "Mandy."

Construction of the marking and targeting plasmids from the above listed DNA elements was carried out following conventional cloning techniques (see, e.g., Molecular Cloning, A Laboratory Manual, J. Sambrook et al, 1987, Cold Spring Harbor Laboratory Press, and Current Protocols in Molecular Biology, F. M. Ausubel et al, eds., 1987, John Wiley and Sons). All plasmids were propagated and maintained in E. coli XLI blue (Stratagene, cat. # 200236). Large scale plasmid preparations were prepared using Promega Wizard Maxiprep DNA Purification System®, according to the manufacturer's directions.

- 31 -

EXAMPLE 2

Construction of a Marked CHO Cell Line

1. Cell Culture and Transfection Procedures to Produced Marked CHO Cell Line

- 5 Marker plasmid DNA was linearized by digestion overnight at 37°C with Bst1107I. Linearized vector was ethanol precipitated and resuspended in sterile TE to a concentration of 1mg/ml. Linearized vector was introduced into DHFR-Chinese hamster ovary cells (CHO cells)
- 10 DG44 cells (Urlaub et al, *Som. Cell and Mol. Gen.*, 12:555-566 (1986)) by electroporation as follows.
- Exponentially growing cells were harvested by centrifugation, washed once in ice cold SBS (sucrose buffered solution, 272mM sucrose, 7mM sodium phosphate,
- 15 pH 7.4, 1mM magnesium chloride) then resuspended in SBS to a concentration of 10^7 cells/ml. After a 15 minute incubation on ice, 0.4ml of the cell suspension was mixed with 40 μ g linearized DNA in a disposable electroporation cuvette. Cells were shocked using a BTX
- 20 electrocell manipulator (San Diego, CA) set at 230 volts, 400 microfaraday capacitance, 13 ohm resistance. Shocked cells were then mixed with 20 ml of prewarmed CHO growth media (CHO-S-SFMII, Gibco/BRL, catalog # 31033-012) and plated in 96 well tissue culture plates.
- 25 Forty eight hours after electroporation, plates were fed with selection media (in the case of transfection with Desmond, selection media is CHO-S-SFMII without

- 32 -

hypoxanthine or thymidine, supplemented with 2mM
Histidinol (Sigma catalog # H6647)). Plates were main-
tained in selection media for up to 30 days, or until
some of the wells exhibited cell growth. These cells
5 were then removed from the 96 well plates and expanded
ultimately to 120 ml spinner flasks where they were
maintained in selection media at all times.

EXAMPLE 3

Characterization of Marked CHO Cell Lines

10 (a) Southern Analysis

Genomic DNA was isolated from all stably growing
Desmond marked CHO cells. DNA was isolated using the
Invitrogen Easy® DNA kit, according to the manufactur-
er's directions. Genomic DNA was then digested with
15 HindIII overnight at 37°C, and subjected to Southern
analysis using a PCR generated digoxigenin labelled
probe specific to the DHFR gene. Hybridizations and
washes were carried out using Boehringer Mannheim's DIG
easy hyb (catalog # 1603 558) and DIG Wash and Block
20 Buffer Set (catalog # 1585 762) according to the manu-
facturer's directions. DNA samples containing a single
band hybridizing to the DHFR probe were assumed to be
Desmond clones arising from a single cell which had
integrated a single copy of the plasmid. These clones
25 were retained for further analysis. Out of a total of
45 HisD resistant cell lines isolated, only 5 were

- 33 -

single copy integrants. Figure 4 shows a Southern blot containing all 5 of these single copy Desmond clones. Clone names are provided in the figure legend.

(b) Northern Analysis

5 Total RNA was isolated from all single copy Desmond clones using TRIzol reagent (Gibco/BRL cat # 15596-026) according to the manufacturer's directions. 10-20 μ g RNA from each clone was analyzed on duplicate formaldehyde gels. The resulting blots were probed with PCR
10 generated digoxigenin labelled DNA probes to (i) DHFR message, (ii) HisD message and (iii) CAD message. CAD is a trifunctional protein involved in uridine biosynthesis (Wahl et al, *J. Biol. Chem.*, 254, 17:8679-8689 (1979)), and is expressed equally in all cell
15 types. It is used here as an internal control to help quantitate RNA loading. Hybridizations and washes were carried out using the above mentioned Boehringer Mannheim reagents. The results of the Northern analysis are shown in Figure 5. The single copy Desmond clone
20 exhibiting the highest levels of both the His D and DHFR message is clone 15C9, shown in lane 4 in both panels of the figure. This clone was designated as the "marked cell line" and used in future targeting experiments in CHO, examples of which are presented in the following
25 sections.

- 34 -

EXAMPLE 4Expression of Anti-CD20 Antibody
in Desmond Marked CHO Cells

C2B8, a chimeric antibody which recognizes B-cell surface antigen CD20, has been cloned and expressed previously in our laboratory. (Reff et al, *Blood*, 83:434-45 (1994)). A 4.1 kb DNA fragment comprising the C2B8 light and heavy chain genes, along with the necessary regulatory elements (eukaryotic promoter and poly-adenylation signals) was inserted into the artificial intron created between exons 1 and 2 of the neo gene contained in a pBR derived cloning vector. This newly generated 5kb DNA fragment (comprising neo exon 1, C2B8 and neo exon 2) was excised and used to assemble the targeting plasmid Molly. The other DNA elements used in the construction of Molly are identical to those used to construct the marking plasmid Desmond, identified previously. A complete map of Molly is shown in Fig. 2.

The targeting vector Molly was linearized prior to transfection by digestion with *Kpn*I and *Pac*I, ethanol precipitated and resuspended in sterile TE to a concentration of 1.5mg/mL. Linearized plasmid was introduced into exponentially growing Desmond marked cells essentially as described, except that 80µg DNA was used in each electroporation. Forty eight hours postelectroporation, 96 well plates were supplemented with selection medium - CHO-SSFMI supplemented with 400 µg/mL Geneti-

- 35 -

cin (G418, Gibco/BRL catalog # 10131-019). Plates were maintained in selection medium for up to 30 days, or until cell growth occurred in some of the wells. Such growth was assumed to be the result of clonal expansion
5 of a single G418 resistant cell. The supernatants from all G418 resistant wells were assayed for C2B8 production by standard ELISA techniques, and all productive clones were eventually expanded to 120mL spinner flasks and further analyzed.

10 Characterization of Antibody secreting Targeted Cells

A total of 50 electroporations with Molly targeting plasmid were carried out in this experiment, each of which was plated into separate 96 well plates. A total of 10 viable, anti-CD20 antibody secreting clones were
15 obtained and expanded to 120ml spinner flasks. Genomic DNA was isolated from all clones, and Southern analyses were subsequently performed to determine whether the clones represented single homologous recombination events or whether additional random integrations had
20 occurred in the same cells. The methods for DNA isolation and Southern hybridization were as described in the previous section. Genomic DNA was digested with EcoRI and probed with a PCR generated digoxigenin labelled probe to a segment of the CD20 heavy chain constant
25 region. The results of this Southern analysis are presented in figure 6. As can be seen in the figure, 8 of

- 36 -

the 10 clones show a single band hybridizing to the CD20 probe, indicating a single homologous recombination event has occurred in these cells. Two of the ten, clones 24G2 and 28C9, show the presence of additional
5 band(s), indicative of an additional random integration elsewhere in the genome.

We examined the expression levels of anti-CD20 antibody in all ten of these clones, the data for which is shown in Table 1, below.

10

Table 1:

Expression Level of Anti-CD20
Secreting Homologous Integrants

| | <u>Clone</u> | <u>Anti-CD20, pg/c/d</u> |
|----|--------------|--------------------------|
| | 20F4 | 3.5 |
| 15 | 25E1 | 2.4 |
| | 42F9 | 1.8 |
| | 39G11 | 1.5 |
| | 21C7 | 1.3 |
| | 50G10 | 0.9 |
| 20 | 29F9 | 0.8 |
| | 5F9 | 0.3 |
| | ----- | |
| | 28C9* | 4.5 |
| | 24G2* | 2.1 |

- 37 -

5 * These clones contained additional randomly integrated copies of anti-CD20. Expression levels of these clones therefore reflect a contribution from both the homologous and random sites.

Expression levels are reported as picogram per cell per day (pg/c/d) secreted by the individual clones, and represented the mean levels obtained from three separate ELISAs on samples taken from 120 mL spinner flasks.

10 As can be seen from the data, there is a variation in antibody secretion of approximately ten fold between the highest and lowest clones. This was somewhat unexpected as we anticipated similar expression levels from all clones due to the fact the anti-CD20 genes are all
15 integrated into the same Desmond marked site. Nevertheless, this observed range in expression extremely small in comparison to that seen using any traditional random integration method or with our translationally impaired vector system.

20 Clone 20F4, the highest producing single copy integrant was selected for further study. Table 2 (below) presents ELISA and cell culture data from seven day production runs of this clone.

- 38 -

Table 2:

7 Day Production Run Data for 20F4

| Day | % Viable | Viable/ml (x 10 ⁵) | Tx2(hr) | mg/L | pg/c/d |
|-----|----------|-----------------------------------|---------|------|--------|
| 1 | 96 | 3.4 | 31 | 1.3 | 4.9 |
| 5 2 | 94 | 6 | 29 | 2.5 | 3.4 |
| 3 | 94 | 9.9 | 33 | 4.7 | 3.2 |
| 4 | 90 | 17.4 | 30 | 6.8 | 3 |
| 5 | 73 | 14 | | 8.3 | |
| 6 | 17 | 3.5 | | 9.5 | |

- 10 Clone 20F4 was seeded at 2×10^5 ml in a 120ml spinner flask on day 0. On the following six days, cell counts were taken, doubling times calculated and 1ml samples of supernatant removed from the flask and analyzed for secreted anti-CD20 by ELISA.
- 15 This clone is secreting on average, 3-5pg antibody/-cell/day, based on this ELISA data. This is the same level as obtained from other high expressing single copy clones obtained previously in our laboratory using the previously developed translationally impaired random
- 20 integration vectors. This result indicates the following:
- (1) that the site in the CHO genome marked by the Desmond marking vector is highly transcriptionally active, and therefore represents an excellent site from
- 25 which to express recombinant proteins, and

- 39 -

(2) that targeting by means of homologous recombination can be accomplished using the subject vectors and occurs at a frequency high enough to make this system a viable and desirable alternative to random integration methods.

To further demonstrate the efficacy of this system, we have also demonstrated that this site is amplifiable, resulting in even higher levels of gene expression and protein secretion. Amplification was achieved by plating serial dilutions of 20F4 cells, starting at a density of 2.5×10^4 cells/ml, in 96 well tissue culture dishes, and culturing these cells in media (CHO-SSFMII) supplemented with 5, 10, 15 or 20nM methotrexate. Antibody secreting clones were screened using standard ELISA techniques, and the highest producing clones were expanded and further analyzed. A summary of this amplification experiment is presented in Table 3 below.

- 40 -

Table 3:

Summary of 20F4 Amplification

| nM MTX | # Wells Assayed | Expression Level mg/l 96 well | # Wells Expanded | Expression Level pg/c/d from spinner |
|--------|-----------------|-------------------------------|------------------|--------------------------------------|
| 10 | 56 | 3-13 | 4 | 10-15 |
| 15 | 27 | 2-14 | 3 | 15-18 |
| 20 | 17 | 4-11 | 1 | ND |

10 Methotrexate amplification of 20F4 was set up as described in the text, using the concentrations of methotrexate indicated in the above table. Supernatants from all surviving 96 well colonies were assayed by ELISA, and the range of anti-CD20 expressed by these clones is indicated in column 3. Based on these results, the highest producing clones were expanded to 120ml spinners and several ELISAs conducted on the spinner supernatants to determine the pg/cell/day expression levels, reported in column 5.

20 The data here clearly demonstrates that this site can be amplified in the presence of methotrexate. Clones from the 10 and 15nM amplifications were found to produce on the order of 15-20pg/cell/day.

A 15nM clone, designated 20F4-15A5, was selected as the highest expressing cell line. This clone originated from a 96 well plate in which only 22 wells grew, and was therefore assumed to have arisen from a single cell.

25 A 15nM clone, designated 20F4-15A5, was selected as the highest expressing cell line. This clone originated

- 41 -

from a 96 well plate in which only 22 wells grew, and was therefore assumed to have arisen from a single cell. The clone was then subjected to a further round of methotrexate amplification. As described above, serial dilutions of the culture were plated into 96 well dishes and cultured in CHO-SS-FMII medium supplemented with 200, 300 or 400nM methotrexate. Surviving clones were screened by ELISA, and several high producing clones were expanded to spinner cultures and further analyzed. A summary of this second amplification experiment is presented in Table 4.

Table 4:
Summary of 20F4-15A5 Amplification

| nM MTX | # Wells Assayed | Expression Level mg/l 96 well | # Wells Expanded | Expression Level pg/c/d, spinner |
|--------|-----------------|-------------------------------|------------------|----------------------------------|
| 200 | 67 | 23-70 | 1 | 50-60 |
| 250 | 86 | 21-70 | 4 | 55-60 |
| 300 | 81 | 15-75 | 3 | 40-50 |

Methotrexate amplifications of 20F4-15A5 were set up and assayed as described in the text. The highest producing wells, the numbers of which are indicated in column 4, were expanded to 120ml spinner flasks. The expression levels of the cell lines derived from these wells is recorded as pg/c/d in column 5.

The highest producing clone came from the 250nM methotrexate amplification. The 250nM clone, 20F4-15A5-250A6 originated from a 96 well plate in which only wells

- 42 -

grew, and therefore is assumed to have arisen from a single cell. Taken together, the data in Tables 3 and 4 strongly indicates that two rounds of methotrexate amplification are sufficient to reach expression levels of 5 60pg/cell/day, which is approaching the maximum secretion capacity of immunoglobulin in mammalian cells (Reff, M.E., *Curr. Opin. Biotech.*, 4:573-576 (1993)). The ability to reach this secretion capacity with just two amplification steps further enhances the utility of 10 this homologous recombination system. Typically, random integration methods require more than two amplification steps to reach this expression level and are generally less reliable in terms of the ease of amplification. Thus, the homologous system offers a more efficient and 15 time saving method of achieving high level gene expression in mammalian cells.

EXAMPLE 5

Expression of Anti-Human CD23 Antibody in Desmond Marked CHO Cells

20 CD23 is low affinity IgE receptor which mediates binding of IgE to B and T lymphocytes (Sutton, B.J., and Gould, H.J., *Nature*, 366:421-428 (1993)). Anti-human CD23 monoclonal antibody 5E8 is a human gamma-1 monoclonal antibody recently cloned and expressed in our 25 laboratory. This antibody is disclosed in commonly

- 43 -

assigned Serial No. 08/803,085, filed on February 20, 1997.

The heavy and light chain genes of 5E8 were cloned into the mammalian expression vector N5KG1, a derivative of the vector NEOSPLA (Barnett et al, in *Antibody Expression and Engineering*, H.Y Yang and T. Imanaka, eds., pp27-40 (1995)) and two modifications were then made to the genes. We have recently observed somewhat higher secretion of immunoglobulin light chains compared to heavy chains in other expression constructs in the laboratory (Reff et al, 1997, unpublished observations). In an attempt to compensate for this deficit, we altered the 5E8 heavy chain gene by the addition of a stronger promoter/enhancer element immediately upstream of the start site. In subsequent steps, a 2.9kb DNA fragment comprising the 5E8 modified light and heavy chain genes was isolated from the N5KG1 vector and inserted into the targeting vector Mandy. Preparation of 5E8-containing Molly and electroporation into Desmond 15C9 CHO cells was essentially as described in the preceding section.

One modification to the previously described protocol was in the type of culture medium used. Desmond marked CHO cells were cultured in protein-free CD-CHO medium (Gibco-BRL, catalog # AS21206) supplemented with 3mg/L recombinant insulin (3mg/mL stock, Gibco-BRL, catalog # AS22057) and 8mM L-glutamine (200mM stock, Gibco-BRL, catalog # 25030-081). Subsequently, trans-

- 44 -

fectected cells were selected in the above medium supplemented with 400 μ g/mL geneticin. In this experiment, 20 electroporations were performed and plated into 96 well tissue culture dishes. Cells grew and secreted anti-
5 CD23 in a total of 68 wells, all of which were assumed to be clones originating from a single G418 cell. Twelve of these wells were expanded to 120ml spinner flasks for further analysis. We believe the increased number of clones isolated in this experiment (68 compared with 10 for anti-CD20 as described in Example 4)
10 is due to a higher cloning efficiency and survival rate of cells grown in CD-CHO medium compared with CHO-SS-FMII medium. Expression levels for those clones analyzed in spinner culture ranged from 0.5-3pg/c/d, in
15 close agreement with the levels seen for the anti-CD20 clones. The highest producing anti-CD23 clone, designated 4H12, was subjected to methotrexate amplification in order to increase its expression levels. This amplification was set up in a manner similar to that described
20 for the anti-CD20 clone in Example 4. Serial dilutions of exponentially growing 4H12 cells were plated into 96 well tissue culture dishes and grown in CD-CHO medium supplemented with 3mg/L insulin, 8mM glutamine and 30, 35 or 40nM methotrexate. A summary of this
25 amplification experiment is presented in Table 5.

Table 5:

- 45 -

Summary of 2H12 Amplification

| nM MTX | # Wells Assayed | Expression Level mg/l 96 well | # Wells Expanded | Expression Level pg/c/d from spinner |
|--------|-----------------|-------------------------------|------------------|--------------------------------------|
| 30 | 100 | 6-24 | 8 | 10-25 |
| 35 | 64 | 4-27 | 2 | 10-15 |
| 5 40 | 96 | 4-20 | 1 | ND |

10 The highest expressing clone obtained was a 30nM clone, isolated from a plate on which 22 wells had grown. This clone, designated 4H12-30G5, was reproducibly secreting 18-22pg antibody per cell per day. This is the same range of expression seen for the first amplification of the anti CD20 clone 20F4 (clone 20F4-15A5 which produced 15-18pg/c/d, as described in Example 4). This data serves to further support the observation that amplification at this marked site in CHO is reproducible and efficient. A second amplification of this 15 30nM cell line is currently underway. It is anticipated that saturation levels of expression will be achievable for the anti-CD23 antibody in just two amplification steps, as was the case for anti-CD20.

20

EXAMPLE 6Expression of Immunoadhesin in Desmond Marked CHO Cells

CTLA-4, a member of the Ig superfamily, is found on the surface of T lymphocytes and is thought to play a role in antigen-specific T-cell activation (Dariavach et al, Eur. J. Immunol., 18:1901-1905 (1988); and Linsley et al, J. Exp. Med., 174:561-569 (1991)). In order to further study the precise role of the CTLA-4 molecule in the activation pathway, a soluble fusion protein comprising the extracellular domain of CTLA-4 linked to a truncated form of the human IgG1 constant region was 30

- 46 -

created (Linsley et al (Id.)). We have recently expressed this CTLA-4 Ig fusion protein in the mammalian expression vector BLECH1, a derivative of the plasmid NEOSPLA (Barnett et al, in Antibody Expression and Engineering, H.Y Yang and T. Imanaka, eds., pp27-40 (1995)).
5 An 800bp fragment encoding the CTLA-4 Ig was isolated from this vector and inserted between the SacII and BglIII sites in Molly.

Preparation of CTLA-4Ig-Molly and electroporation
10 into Desmond clone 15C9 CHO cells was performed as described in the previous example relating to anti-CD20. Twenty electroporations were carried out, and plated into 96 well culture dishes as described previously. Eighteen CTLA-4 expressing wells were isolated from the
15 96 well plates and carried forward to the 120ml spinner stage. Southern analyses on genomic DNA isolated from each of these clones were then carried out to determine how many of the homologous clones contained additional random integrants. Genomic DNA was digested with BglIII
20 and probed with a PCR generated digoxigenin labelled probe to the human IgG1 constant region. The results of this analysis indicated that 85% of the CTLA-4 clones are homologous integrants only; the remaining 15% contained one additional random integrant. This result
25 corroborates the findings from the expression of anti-CD20 discussed above, where 80% of the clones were single homologous integrants. Therefore, we can conclude

- 47 -

that this expression system reproducibly yields single targeted homologous integrants in at least 80% of all clones produced.

Expression levels for the homologous CTLA4-Ig clones ranged from 8-12pg/cell/day. This is somewhat higher than the range reported for anti-CD20 antibody and anti-CD23 antibody clones discussed above. However, we have previously observed that expression of this molecule using the intronic insertion vector system also resulted in significantly higher expression levels than are obtained for immunoglobulins. We are currently unable to provide an explanation for this observation.

EXAMPLE 7

Targeting Anti-CD20 to an alternate Desmond Marked CHO Cell Line

As we described in a preceding section, we obtained 5 single copy Desmond marked CHO cell lines (see Figures 4 and 5). In order to demonstrate that the success of our targeting strategy is not due to some unique property of Desmond clone 15C9 and limited only to this clone, we introduced anti-CD20 Molly into Desmond clone 9B2 (lane 6 in figure 4, lane 1 in figure 5). Preparation of Molly DNA and electroporation into Desmond 9B2 was exactly as described in the previous example pertaining to anti-CD20. We obtained one homologous integrant from this experiment. This clone was expanded to a 120ml

- 48 -

spinner flask, where it produced on average 1.2pg anti-
CD20/cell/day. This is considerably lower expression
than we observed with Molly targeted into Desmond 15C9.
However, this was the anticipated result, based on our
5 northern analysis of the Desmond clones. As can be seen
in Figure 5, mRNA levels from clone 9B2 are considerably
lower than those from 15C9, indicating the site in this
clone is not as transcriptionally active as that in
15C9. Therefore, this experiment not only demonstrates
10 the reproducibility of the system - presumably any
marked Desmond site can be targeted with Molly - it also
confirms the northern data that the site in Desmond 15C9
is the most transcriptionally active.

From the foregoing, it will be appreciated that,
15 although specific embodiments of the invention have been
described herein for purposes of illustration, various
modifications may be made without diverting from the
scope of the invention. Accordingly, the invention is
not limited by the appended claims.

- 49 -

WHAT IS CLAIMED IS:

1. A method for inserting a desired DNA at a target site in the genome of a mammalian cell which comprises the following steps:

5 (i) transfecting or transforming a mammalian cell with a first plasmid ("marker plasmid") containing the following sequences:

(a) a region of DNA that is heterologous to the mammalian cell genome which when integrated in the
10 mammalian cell genome provides a unique site for homologous recombination;

(b) a DNA fragment encoding a portion of a first selectable marker protein; and

(c) at least one other selectable marker DNA
15 that provides for selection of mammalian cells which have been successfully integrated with the marker plasmid;

(ii) selecting a cell which contain the marker plasmid integrated in its genome;

20 (iii) transfecting or transforming said selected cell with a second plasmid ("target plasmid") which contains the following sequences:

(a) a region of DNA that is identical or is sufficiently homologous to the unique region in the
25 marker plasmid such that this region of DNA can recombine with said DNA via homologous recombination;

- 50 -

- (b) a DNA fragment encoding a portion of the same selectable marker contained in the marker plasmid, wherein the active selectable marker protein encoded by said DNA is only produced if said fragment is expressed
5 in association with the fragment of said selectable marker DNA contained in the marker plasmid; and
- (iv) selecting cells which contain the target plasmid integrated at the target site by screening for the expression of the first selectable marker protein.

10 2. The method of Claim 1, wherein the DNA fragment encoding a fragment of a first selectable marker is an exon of a dominant selectable marker.

3. The method of Claim 2, wherein the second plasmid contains the remaining exons of said first
15 selectable marker.

4. The method of Claim 3, wherein at least one DNA encoding a desired protein is inserted between said exons of said first selectable marker contained in the target plasmid.

20 5. The method Claim 4, wherein a DNA encoding a dominant selectable marker is further inserted between the exons of said first selectable marker contained in

- 51 -

the target plasmid to provide for co-amplification of the DNA encoding the desired protein.

6. The method of Claim 3, wherein the first dominant selectable marker is selected from the group consisting of neomycin phosphotransferase, histidinol dehydrogenase, dihydrofolate reductase, hygromycin phosphotransferase, herpes simplex virus thymidine kinase, adenosine deaminase, glutamine synthetase, and hypoxanthine-guanine phosphoribosyl transferase.

10 7. The method of Claim 4, wherein the desired protein is a mammalian protein.

8. The method of Claim 7, wherein the protein is an immunoglobulin.

15 9. The method of Claim 1, which further comprises determining the RNA levels of the selectable marker (c) contained in the marker plasmid prior to integration of the target vector.

20 10. The method of Claim 9, wherein the other selectable marker contained in the marker plasmid is a dominant selectable marker selected from the group consisting of histidinol dehydrogenase, herpes simplex

- 52 -

thymidine kinase, hydromycin phosphotransferase, adenosine deaminase and glutamine synthetase.

11. The method of Claim 1, wherein the mammalian cell is selected from the group consisting of Chinese hamster ovary (CHO) cells, myeloma cells, baby hamster kidney cells, COS cells, NSO cells, HeLa cells and NIH 3T3 cells.

12. The method of Claim 11, wherein the cell is a CHO cell.

10 13. The method of Claim 1, wherein the marker plasmid contains the third exon of the neomycin phosphotransferase gene and the target plasmid contains the first two exons of the neomycin phosphotransferase gene.

14. The method of Claim 1, wherein the marker plasmid further contains a rare restriction endonuclease sequence which is inserted within the region of homology.

15 15. The method of Claim 1, wherein the unique region of DNA that provides for homologous recombination is a bacterial DNA, a viral DNA or a synthetic DNA.

20

- 53 -

16. The method of Claim 1, wherein the unique region of DNA that provides for homologous recombination is at least 300 nucleotides.

5 17. The method of Claim 16, wherein the unique region of DNA ranges in size from about 300 nucleotides to 20 kilobases.

18. The method of claim 17, wherein the unique region of DNA preferably ranges in size from 2 to 10 kilobases.

10 19. The method of Claim 1, wherein the first selectable marker DNA is split into at least three exons.

20. The method of Claim 1, wherein the unique region of DNA that provides for homologous recombination
15 is a bacterial DNA, an insect DNA, a viral DNA or a synthetic DNA.

21. The method of Claim 20, wherein the unique region of DNA does not contain any functional genes.

22. A vector system for inserting a desired DNA at
20 a target site in the genome of a mammalian cell which comprises at least the following:

- 54 -

(i) a first plasmid ("marker plasmid") containing at least the following sequences:

(a) a region of DNA that is heterologous to the mammalian cell genome which when integrated in the mammalian cell genome provides a unique site for homologous recombination;

(b) a DNA fragment encoding a portion of a first selectable marker protein; and

(c) at least one other selectable marker DNA that provides for selection of mammalian cells which have been successfully integrated with the marker plasmid; and

(ii) a second plasmid ("target plasmid") which contains at least the following sequences:

(a) a region of DNA that is identical or is sufficiently homologous to the unique region in the marker plasmid such that this region of DNA can recombine with said DNA via homologous recombination;

(b) a DNA fragment encoding a portion of the same selectable marker contained in the marker plasmid, wherein the active selectable marker protein encoded by said DNA is only produced if said fragment is expressed in association with the fragment of said selectable marker DNA contained in the marker plasmid.

- 55 -

23. The vector system of Claim 22, wherein the DNA fragment encoding a fragment of a first selectable marker is an exon of a dominant selectable marker.

24. The vector system of Claim 23, wherein the
5 second plasmid contains the remaining exons of said first selectable marker.

25. The vector system of Claim 24, wherein at least one DNA encoding a desired protein is inserted between said exons of said first selectable marker con-
10 tained in the target plasmid.

26. The vector system of Claim 24, wherein a DNA encoding a dominant selectable marker is further inserted between the exons of said first selectable marker contained in the target plasmid to provide for co-ampli-
15 fication of the DNA encoding the desired protein.

27. The vector system of Claim 24, wherein the first dominant selectable marker is selected from the group consisting of neomycin phosphotransferase, histidinol dehydrogenase, dihydrofolate reductase, hygromycin phosphotransferase, herpes simplex virus
20 thymidine kinase, adenosine deaminase, glutamine synthetase, and hypoxanthine-guanine phosphoribosyl transferase.

- 56 -

28. The vector system of Claim 25, wherein the desired protein is a mammalian protein.

29. The vector system of Claim 28, wherein the protein is an immunoglobulin.

5 30. The vector system of Claim 22, wherein the other selectable marker contained in the marker plasmid is a dominant selectable marker selected from the group consisting of histidinol dehydrogenase, herpes simplex thymidine kinase, hydromycin phosphotransferase, adeno-
10 sine deaminase and glutamine synthetase.

31. The vector system of Claim 22, which provides for insertion of a desired DNA at a targeted site in the genome of a mammalian cell selected from the group consisting of Chinese hamster ovary (CHO) cells, myeloma
15 cells, baby hamster kidney cells, COS cells, NSO cells, HeLa cells and NIH 3T3 cells.

32. The vector system of Claim 31, wherein the mammalian cell is a CHO cell.

33. The vector system of Claim 22, wherein the
20 marker plasmid contains the third exon of the neomycin phosphotransferase gene and the target plasmid contains

- 57 -

the first two exons of the neomycin phosphotransferase gene.

34. The vector system of Claim 22, wherein the marker plasmid further contains a rare restriction endo-
5 nuclease sequence which is inserted within the region of homology.

35. The vector system of Claim 22, wherein the unique region of DNA that provides for homologous recombination is a bacterial DNA, a viral DNA or a synthetic
10 DNA.

36. The vector system of Claim 22, wherein the unique region of DNA (a) contained in the marker plasmid vector system that provides for homologous recombination is at least 300 nucleotides.

15 37. The vector system of Claim 36, wherein the unique region of DNA ranges in size from about 300 nucleotides to 20 kilobases.

38. The vector system of Claim 37, wherein the unique region of DNA preferably ranges in size from 2 to
20 10 kilobases.

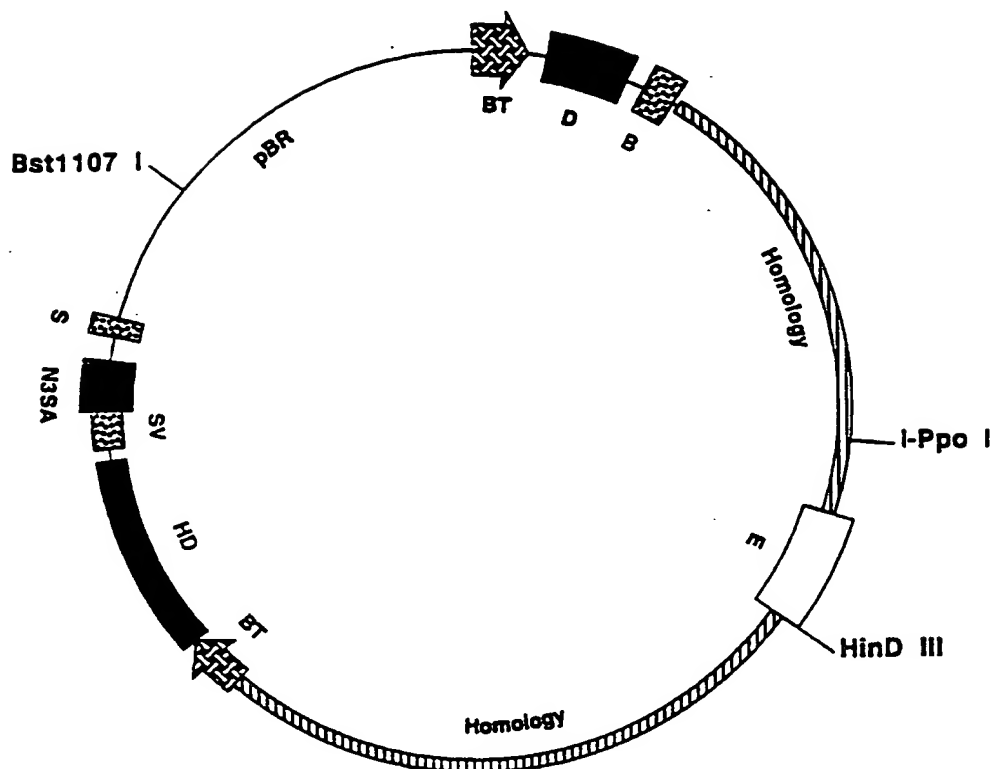
- 58 -

39. The vector system of Claim 22, wherein the first selectable marker DNA is split into at least three exons.

40. The vector system of Claim 22, wherein the
5 unique region of DNA that provides for homologous recombination is a bacterial DNA, an insect DNA, a viral DNA or a synthetic DNA.

41. The vector system of Claim 40, wherein the
10 unique region of DNA does not contain any functional genes.

DESMOND



- HD = Salmonella HisD Gene
N3 = Neomycin Phosphotransferase Exon 3
D = Murine Dihydrofolate reductase
E = Cytomegalovirus and SV40 Enhancers
SA = Splice acceptor
BT = Mouse Beta Globin Major Promoter
B = Bovine Growth Hormone Polyadenylation
S = SV40 Early Polyadenylation
SV = SV40 Late Polyadenylation

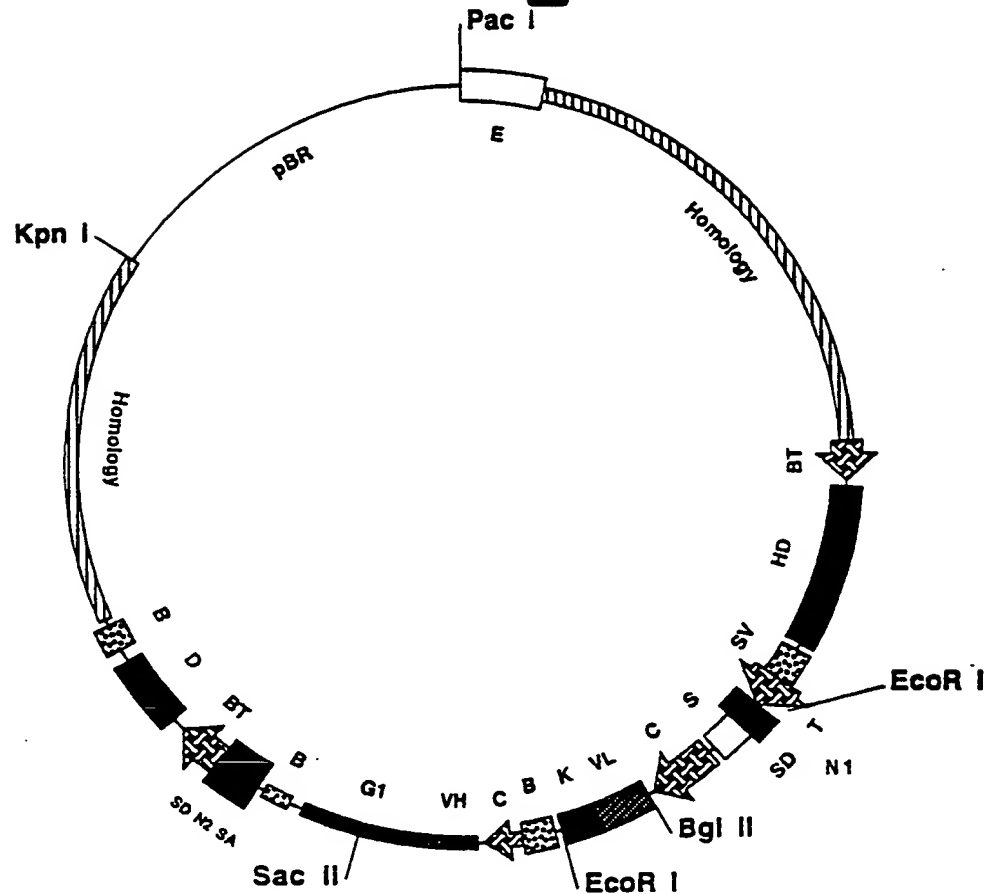
FIGURE 1A

Desmond
14,683 bp Bst1107 I linear



FIGURE 1B

Molly



- D = Dihydrofolate reductase
 N1 = Neomycin Phosphotransferase Exon 1
 N2 = Neomycin Phosphotransferase Exon 2
 VL = Anti-CD20 Light chain leader + Variable
 K = Human Kappa Constant
 VH = Anti-CD20 Heavy chain Leader + Variable
 G1 = Human Gamma 1 Constant
 HD = Salmonella Histidinol Dehydrogenase
 E = CMV and SV40 enhancers S = SV40 Origin
 SD = Splice donor SA = Splice acceptor
 C = CMV promoter/enhancer
 T = HSV TK promoter and Polyoma enhancers
 BT = Mouse Beta Globin Major Promoter
 SV = SV40 Late Polyadenylation
 B = Bovine Growth Hormone Polyadenylation

FIGURE 2A

Molly
15,987 bp Pac I, Kpn I fragment

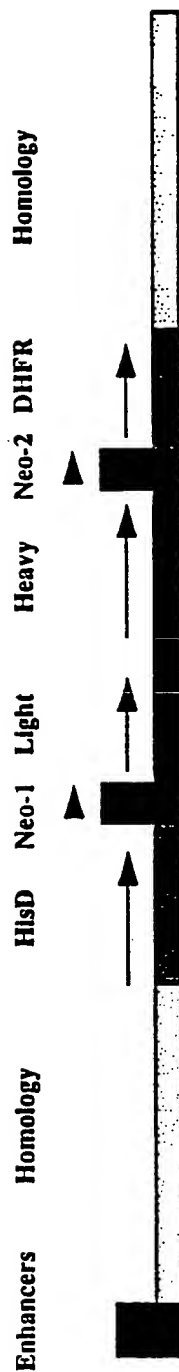


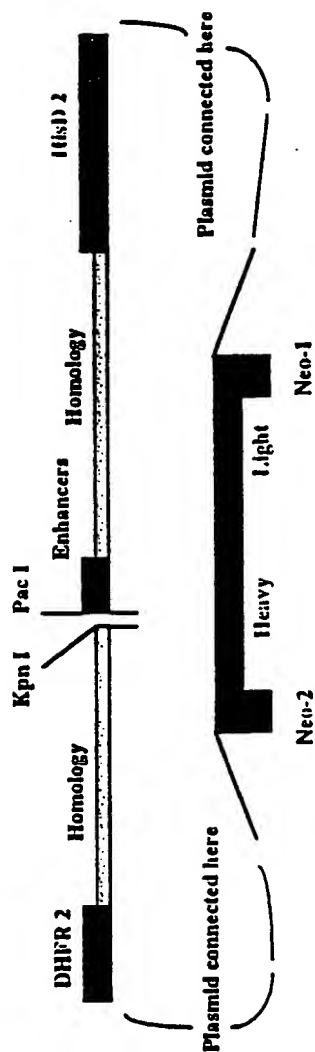
FIGURE 2B

Homologous Recombination

Desmond in CHO



Molly



Single crossover in CHO



FIGURE 7

Southern Analysis of Desmond Marked CHO Cells

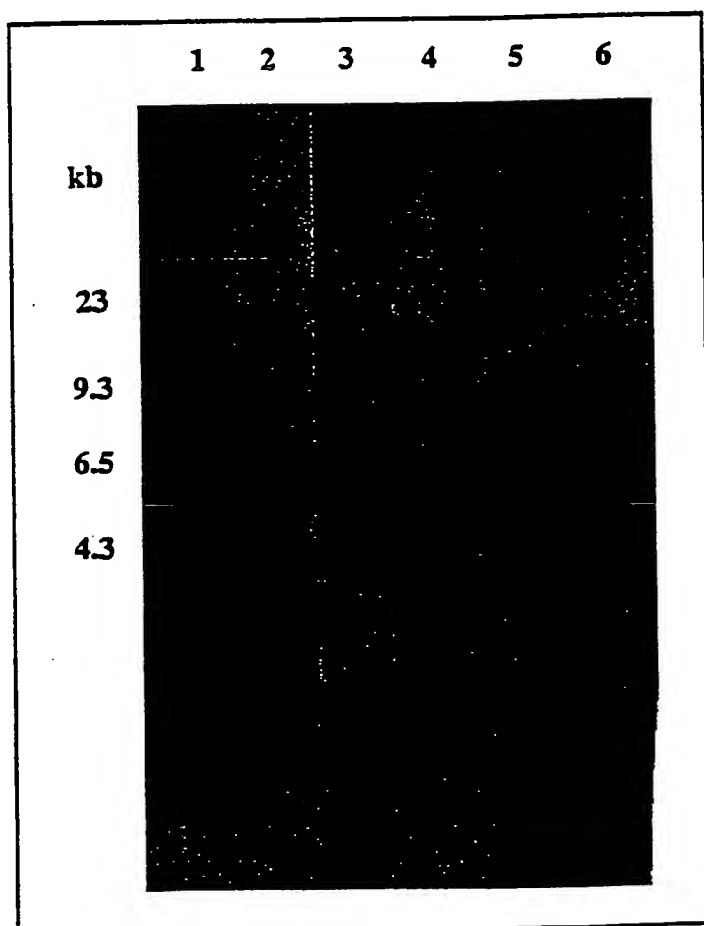


FIGURE 4

Northern Analysis of Desmond
Marked CHO Cells

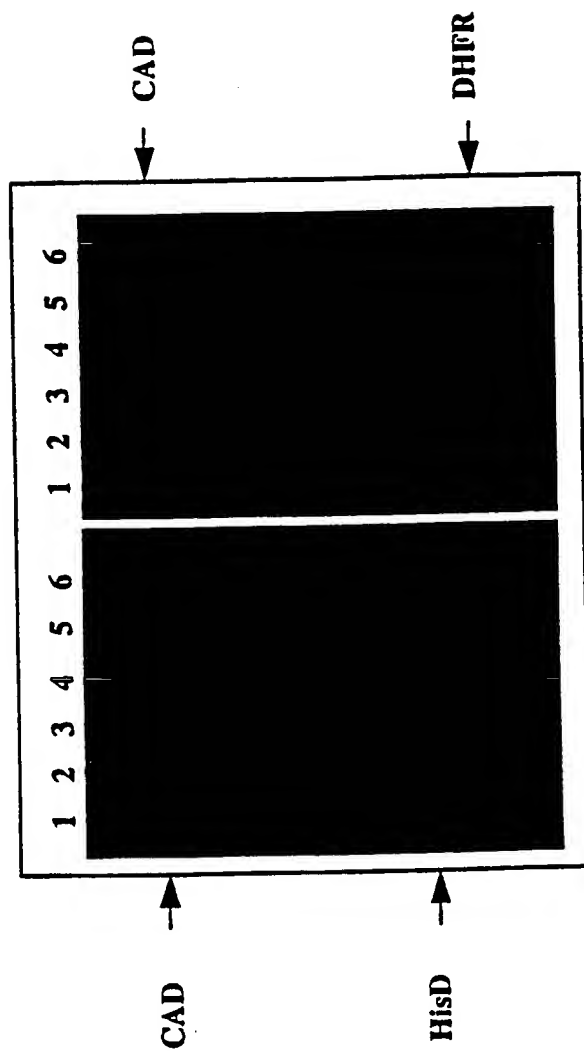


FIGURE 5

Southern Analysis of Anti CD20
Integrants in Marked CHO Cells

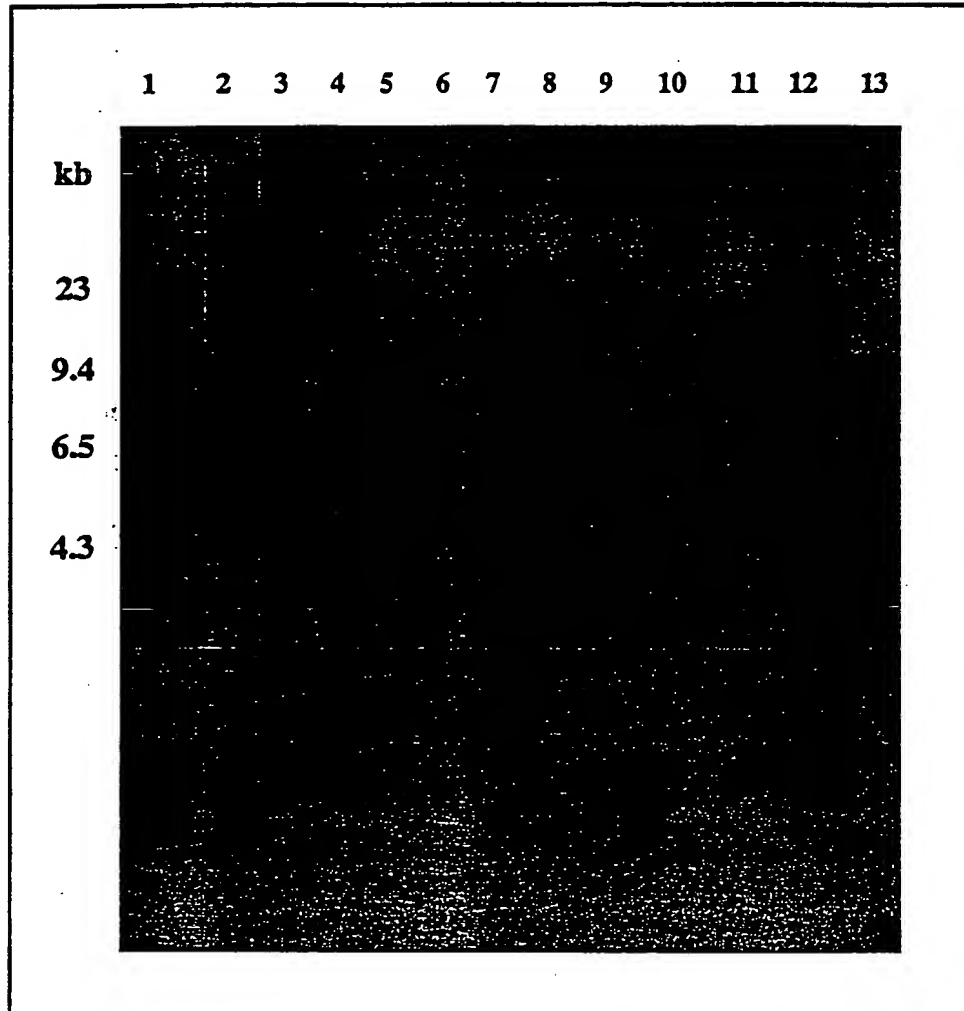


FIGURE 6

DNASIS
Desmond Lark

```

      10      20      30      40      50      60
TTTCTAGACC TAGGGCGGCC AGCTAGTAGC TTTGCTTCTC AATTCTTAT TTGCATAATG

      70      80      90     100     110     120
AGAAAAAAG GAAAATTAAT TTAAACACCA ATTCAGTAGT TGATTGAGCA AATGCGTTGC

      130     140     150     160     170     180
CAAAAAGGAT GCTTTAGAGA CAGTGTCTC TGCACAGATA AGGACAAACA TTATTCAGAG

      190     200     210     220     230     240
GGAGTACCCA GAGCTGAGAC TCCTAAGCCA GTGAGTGGCA CAGCATCCAG GGAGAAATAT

      250     260     270     280     290     300
GCTTGTCTAT ACCGAAGCCT GATTCCGTAG AGCCACACCC TGGTAAGGGC CAATCTGCTC

      310     320     330     340     350     360
ACACAGGATA GAGAGGGCAG GAGCCAGGGC AGAGCATATA AGGTGAGGTA GGATCAGTTG

      370     380     390     400     410     420
CTCATAT TTGCTTCTGA CATAGTTGTG TTGGGAGCTT GGATAGCTTG GGGGGGGGAC

      430     440     450     460     470     480
AGCTCAGGGC TGGGATTTCG CGCCAACTT GACGGCAATC CTAGCGTGAA GGCTGGTAGG

      490     500     510     520     530     540
ATTTTATCCC CGCTGCCATC ATGGTTCGAC CATTGAACTG CATCGTCGCC GTGTCCCAAA

      550     560     570     580     590     600
ATATGGGGAT TGGCAAGAAC GGAGACCTAC CCTGGCCTCC GCTCAGGAAC GAGTTCAAGT

      610     620     630     640     650     660
ACTTCCAAAG AATGACCACA ACCTCTTCAG TGGGAAGGTAA ACAGAATCTG GTGATTATGG

      670     680     690     700     710     720
GTAGGAAAC CTGGTCTCC ATTCCTGAGA AGAATCGACC TTTAAAGGAC AGAATTAATA

      730     740     750     760     770     780
TTCTCAG TAGAGAACTC AAAGAACCAC CACGAGGAGC TCATTTTCTT GCCAAAAGTT

      790     800     810     820     830     840
TGGATGATGC CTTAAGACTT ATTGAACAAC CGGAATTGGC AAGTAAAGTA GACATGGTTT

      850     860     870     880     890     900
GGATAGTCGG AGGCAGTTCT GTTTACCAGG AAGCCATGAA TCAACCAGGC CACCTCAGAC

      910     920     930     940     950     960
TCTTTGTGAC AAGGATCATG CAGGAATTTG AAAGTGACAC GTTTTCCCA GAAATTGATT

      970     980     990    1000    1010    1020
TGGGGAAATA TAACTTCTC CCAGAATACC CAGGCGTCTC CTCTGAGGTC CAGGAGGAAA

     1030    1040    1050    1060    1070    1080
AAGGCATCAA GTATAAGTTT GAAGCTACG AGAAGAAAGA CTAACAGGAA GATGCTTTCA

     1090    1100    1110    1120    1130    1140
AGTTCTCTGC TCCCCTCCTA AAGCTATGCA TTTTATAAG ACCATGGGAC TTTTGCTGGC

     1150    1160    1170    1180    1190    1200
TTTAGATCAG CCTCGACTGT GCCTTCTAGT TGCCAGCCAT CTGTTGTTTG CCCCTCCCCC

     1210    1220    1230    1240    1250    1260
GTGCCTTCTC TGACCCTGGA AGGTGCCACT CCCACTGTCC TTTCCTAATA AAATGAGGAA

     1270    1280    1290    1300    1310    1320
ATTGCATCGC ATTGTCTGAG TAGGTGTCAT TCTATTCTGG GGGGTGGGGT GGGGCAGGAC

```

FIGURE 7

DNA 515
Desmond Lark

```

1330      1340      1350      1360      1370      1380
AGCAAGGGGG AGGATTGGGA AGACAATAGC AGGCATGCTG GGGATGCCGT GGGCTCTATG

1390      1400      1410      1420      1430      1440
GCTTCTGAGG CGGAAAGAAC CAGCTGGGGC TCGAAGCGGC CGCCCATTTT GCTGGTGGTC

1450      1460      1470      1480      1490      1500
AGATGCCGGA TGGCGTGGGA CGCGGGGGGG AGCGTCACAC TGAGGTTTTT CGCCAGACGC

1510      1520      1530      1540      1550      1560
CACTGCTGCC AGGCGCTGAT GTGCCCCGGT TCTGACCATG CGGTCGCGTT CGGTTGCACT

1570      1580      1590      1600      1610      1620
ACGCGTACTG TGAGCCAGAG TTGCCCCGGC CTCTCCGGCT GCGGTAGTTC AGGCAGTTCA

1630      1640      1650      1660      1670      1680
ATCAACTGTT TACCTTGTTG AGCGACATCC AGAGGCACTT CACCGCTTGC CAGCGGCTTA

1690      1700      1710      1720      1730      1740
ATCCAGCG CCACCATCCA GTGCAGGAGC TCGTTATCGC TATGACGGAA CAGGTATTGC

1750      1760      1770      1780      1790      1800
CTGGTCACTT CGATGGTTTG CCCGGATAAA CGGAACTGGA AAAACTGCTG CTGGTGTTTT

1810      1820      1830      1840      1850      1860
GCTTCGTC GCGCTGGATG CGGCGTGGG TCGGCAAAGA CCAGACCGTT CATACAGAAC

1870      1880      1890      1900      1910      1920
TGCGGATCGT TCGGCGTATC GCCAAAATCA CCGCCGTAAG CCGACCACGG GTTGCCGTTT

1930      1940      1950      1960      1970      1980
TCATCATATT TAATCAGCGA CTGATCCACC CAGTCCCAGA CGAAGCCGCC CTGTAAACGG

1990      2000      2010      2020      2030      2040
GGATACTGAC GAAACGCCTG CCAGTATTTA GCGAAACCGC CAAGACTGTT ACCCATCGCG

2050      2060      2070      2080      2090      2100
GGCGTATT CGCAAAGGAT CAGCGGGCGC GTCTCTCCAG GTAGCGAAAG CCATTTTTTG

2110      2120      2130      2140      2150      2160
ATGGACCATT TCGGCACAGC CGGGAAGGGC TGGTCTTCAT CCACGCGCGC GTACATCGGG

2170      2180      2190      2200      2210      2220
CAAATAATAT CGGTGGCCGT GGTGTCGGCT CCGCCGCCCT CATACTGCAC CGGGCGGGAA

2230      2240      2250      2260      2270      2280
GGATCGACAG ATTTGATCCA GCGATACAGC GCGTCGTGAT TAGCGCCGTG GCCTGATTCA

2290      2300      2310      2320      2330      2340
TTCCCCAGCG ACCAGATGAT CACACTCGGG TGATTACGAT CGCGCTGCAC CATTGCGGTT

2350      2360      2370      2380      2390      2400
ACGCGTTCCG TCATCGCCGG TAGCCAGCGC GGATCATCGG TCAGACGATT CATTGGCACC

2410      2420      2430      2440      2450      2460
ATGCCGTGGG TTTCAATATT GGCTTCATCC ACCACATACA GGCCGTAGCG GTCGCACAGC

2470      2480      2490      2500      2510      2520
GTGTACCACA GCGGATGGTT CGGATAATGC GAACAGCGCA CGGCGTTAAA GTTGTCTGTC

2530      2540      2550      2560      2570      2580
TTCATCAGCA GGATATCCTG CACCATCGTC TGCTCATCCA TGACCTGACC ATGCAGAGGA

2590      2600      2610      2620      2630      2640

```

DNASIS

Desmond Lark

TGATGCTCGT GACGGTTAAC GCCTCGAATC AGCAACGGCT TGCCGTTCAG CAGCAGCAGA
2650 2660 2670 2680 2690 2700
CCATTTTCAA TCCGCACCTC GCGGAAACCG ACATCGCAGG CTTCTGCTTC AATCAGCGTG
2710 2720 2730 2740 2750 2760
CCGTCGGCGG TGTGCAGTTC AACCACCGCA CGATAGAGAT TCGGGATTTC GCGGCTCCAC
2770 2780 2790 2800 2810 2820
AGTTTCGGGT TTTCGACGTT CAGACGTAGT GTGACGCGAT CGGCATAACC ACCACGCTCA
2830 2840 2850 2860 2870 2880
TCGATAATTT CACCGCCGAA AGGCGCGGTG CCGCTGGCGA CCTGCGTTTC ACCCTGCCAT
2890 2900 2910 2920 2930 2940
AAAGAAACTG TTACCCGTAG GTAGTCACGC AACTCGCCGC ACATCTGAAC TTCAGCCTCC
2950 2960 2970 2980 2990 3000
AGTACAGCGC GGCTGAAATC ATCATTAAAG CGAGTGGCAA CATGGAAATC GCTGATTGT
3010 3020 3030 3040 3050 3060
GAGTCGGTT TATGCAGCAA CGAGACGTCA CGGAAAATGC CGCTCATCCG CCACATATCC
3070 3080 3090 3100 3110 3120
TGATCTTCCA GATAACTGCC GTCACCTCAG CGCAGCACCA TCACCGCGAG GCGGTTTTCT
3130 3140 3150 3160 3170 3180
CCGGCGCGTA AAAATGCGCT CAGGTCAAAT TCAGACGGCA AACGACTGTC CTGGCCGTAA
3190 3200 3210 3220 3230 3240
CCGACCCAGC GCGCGTTGCA CCACAGATGA AACGCCGAGT TAACGCCATC AAAAATAATT
3250 3260 3270 3280 3290 3300
CGGCTCTGGC CTTCTGTAG CCAGCTTTCA TCAACATTAA ATGTGAGCGA GTAACAACCC
3310 3320 3330 3340 3350 3360
GTCGGATTCT CCGTGGGAAC AAACGGCGGA TTGACCGTAA TGGGATAGGT CACGTTGGTG
3370 3380 3390 3400 3410 3420
IAGATGGGCG CATCGTAACC GTGCATCTGC CAGTTTGAGG GGACGACGAC AGTATCGGCC
3430 3440 3450 3460 3470 3480
TCAGGAAGAT CGCACTCCAG CCAGCTTTCC GGCACCGCTT CTGGTGCCGG AAACCAGGCA
3490 3500 3510 3520 3530 3540
AAGCGCCATT CGCCATTGAG GCTGCGCAAC TGTGGGAAG GCGGATCGGT GCGGGCCTCT
3550 3560 3570 3580 3590 3600
TCGCTATTAC GCCAGCTGGC GAAAGGGGGA TGTGCTGCAA GGCGATTAAG TTGGGTAACG
3610 3620 3630 3640 3650 3660
CCAGGGTTTT CCCAGTCACG ACGTTGTAAA ACGACTTAAT CCGTCGAGGG GCTGCCTCGA
3670 3680 3690 3700 3710 3720
AGCAAACGAC CTTCCGTTGT GCAGCCAGCG GCGCCTGCGC CGGTGCCCAC AATCGTGCGC
3730 3740 3750 3760 3770 3780
GAACAACTA AACCAGAAC AATTATACCG GCGGCACCGC CGCCACCACC TTCTCCCGTG
3790 3800 3810 3820 3830 3840
CCTAACATTC CAGCGCCTCC ACCACCACCA CCACCATCGA TGTCTGAATT GCCGCCCGCT
3850 3860 3870 3880 3890 3900
CCACCAATGC CGACGGAACC TCAACCCGCT GCACCTTTAG ACGACAGACA ACAATTGTTG

DNASIS
Desmond Lark

3910 3920 3930 3940 3950 3960
GAAGCTATTA GAAACGAAAA AAATCGCACT CGTCTCAGAC CGGCTCTCTT AAGGTAGCTC

3970 3980 3990 4000 4010 4020
AAACCAAAAA CGGCGCCCGA AACCAGTACA ATAGTTGAGG TGCCGACTGT GTTGCCATAA

4030 4040 4050 4060 4070 4080
GAGACATTTG AGCCTAAACC GCCGTCTGCA TCACCGCCAC CACCTCCGCC TCCGCCTCCG

4090 4100 4110 4120 4130 4140
CCGCCAGCCC CGCCTGCGCC TCCACCGATG GTAGATTTAT CATCAGCTCC ACCACGCGCG

4150 4160 4170 4180 4190 4200
CCATTAGTAG ATTTGCCGTC TGAAATGTTA CCACCGCCTG CACCATCGCT TTCTAACGTG

4210 4220 4230 4240 4250 4260
TTGTCTGAAT TAAAATCGGG CACAGTTAGA TTGAAACCCG CCCAAAAACG CCCGCAATCA

4270 4280 4290 4300 4310 4320
AATAATTC CAAAAGCTC AACTACAAAT TTGATCGCGG ACGTGTTAGC CGACACAATT

4330 4340 4350 4360 4370 4380
AATAGGCGTC GTGTGGCTAT GGCAAAATCG TCTTCGGAAG CAACTTCTAA CGACGAGGGT

4390 4400 4410 4420 4430 4440
TGGGACGACG ACGATAATCG GCCTAATAAA GCTAACACGC CCGATGTTAA ATATGTCCAA

4450 4460 4470 4480 4490 4500
GCTACTAGTG GTACCTTAAT TAAGGGGCGG AGAATGGGCG GAACTGGGCG GAGTTAGGGG

4510 4520 4530 4540 4550 4560
CGGGATGGGC GGAGTTAGGG GCGGGACTAT GGTGCTGAC TAATTGAGAT GCATGCTTTG

4570 4580 4590 4600 4610 4620
CATACTTCTG CCTGCTGGGG AGCCTGGGGA CTTCCACAC CTGGTTGCTG ACTAATTGAG

4630 4640 4650 4660 4670 4680
TGATGCTT TGCATACTTC TGCCTGCTGG GGAGCCTGGG GACTTTCCAC ACCCTAACTG

4690 4700 4710 4720 4730 4740
ACACACATTC CACAGAATTA ATTCCCCTAG TTATTAATAG TAATCAATTA CGGGGTCATT

4750 4760 4770 4780 4790 4800
AGTTCATAGC CCATATATGG AGTTCGCGT TACATAACTT ACGGTAAATG GCCCGCCTGG

4810 4820 4830 4840 4850 4860
CTGACCGCCC AACGACCCCC GCCCATTGAC GTCAATAATG ACGTATGTTT CCATAGTAAC

4870 4880 4890 4900 4910 4920
GCCAATAGGG ACTTTCCATT GACGTCAATG GGTGGAGTAT TTACGGTAAA CTGCCCCATT

4930 4940 4950 4960 4970 4980
GGCAGTACAT CAAGTGTATC ATATGCCAAG TACGCCCCCT ATTGACGTCA ATGACGGTAA

4990 5000 5010 5020 5030 5040
ATGGCCCCGC TGGCATTATG CCCAGTACAT GACCTTATGG GACTTTCCTA CTTGGCAGTA

5050 5060 5070 5080 5090 5100
CATCTACGTA TTAGTCATCG CTATTACCAT GGTGATGCGG TTTTGGCAGT ACATCAATGG

5110 5120 5130 5140 5150 5160
GCGTGGATAG CGGTTTGACT CACGGGGATT TCCAAGTCTC CACCCATTG ACGTCAATGG

5170 5180 5190 5200 5210 5220
GAGTTTGTTT TGAAGCTTGG CCGGCCAGCT TTATTTAACG TGTTTACGTC GAGTCAATTG

DNASIS
Desmond Lark

```

5230      5240      5250      5260      5270      5280
TACACTAACG ACAGTGATGA AAGAAATACA AAAGCGCATA ATATTTTGAA CGACGTGAA

5290      5300      5310      5320      5330      5340
CCTTTATTAC AAAACAAAAC ACAAACGAAT ATCGACAAAG CTAGATTGCT GCTACAAGAT

5350      5360      5370      5380      5390      5400
TTGGCAAGTT TTGTGGCGTT GAGCGAAAAT CCATTAGATA GTCCAGCCAT CGGTTGCGAA

5410      5420      5430      5440      5450      5460
AAACAACCCT TGTGTGAAAC TAATCGAAAC CTATTTTACA AATCTATTGA GGATTTAATA

5470      5480      5490      5500      5510      5520
TTTAAATTCA GATATAAAGA CGCTGAAAAT CATTTGATTT TCGCTCTAAC ATACCACCCT

5530      5540      5550      5560      5570      5580
AAAGATTATA AATTTAATGA ATTATTAATA TACATCAGCA ACTATATATT GATAGACATT

5590      5600      5610      5620      5630      5640
CAGTTTGT GATATTAGTT TGTGCGTCTC ATTACAATGG CTGTTATTTT TAACAACAAA

5650      5660      5670      5680      5690      5700
CAACTGCTCG CAGACAATAG TATAGAAAAG GGAGGTGAAC TGTTTTTGT TAACGGTTCG

5710      5720      5730      5740      5750      5760
TACAACATTT TGGAAAGTTA TGTTAATCCG GTGCTGCTAA AAAATGGTGT AATTGAACTA

5770      5780      5790      5800      5810      5820
GAAGAAGCTG CGTACTATGC CGGCAACATA TTGTACAAAA CCGACGATCC CAAATTCATT

5830      5840      5850      5860      5870      5880
GATTATATAA ATTTAATAAT TAAAGCAACA CACTCCGAAG AACTACCAGA AAATAGCACT

5890      5900      5910      5920      5930      5940
GTTGTAAATT ACAGAAAAAC TATGCGCAGC GGTACTATAC ACCCCATTAA AAAAGACATA

5950      5960      5970      5980      5990      6000
...TATTATG ACAACAAAAA ATTTACTCTA TACGATAGAT ACATATATGG ATACGATAAT

6010      6020      6030      6040      6050      6060
AACTATGTTA ATTTTATGA GGAGAAAAAT GAAAAAGAGA AGGAATACGA AGAAGAAGAC

6070      6080      6090      6100      6110      6120
GACAAGGCGT CTAGTTTATG TGAAAAATAA ATTATATTGT CGCAAATTAA CTGTGAATCA

6130      6140      6150      6160      6170      6180
TTTGAAATG ATTTTAAATA TTACCTCAGC GATTATAACT ACGCGTTTTC AATTATAGAT

6190      6200      6210      6220      6230      6240
AATACTACAA ATGTTCTTGT TGCCTTTGGT TTGTATCGTT AATAAAAAAC AAATTTGACA

6250      6260      6270      6280      6290      6300
TTTATAATTG TTTTATTATT CAATAATTAC AAATAGGATT GAGACCCTTG CAGTTGCCAG

6310      6320      6330      6340      6350      6360
CAAACGGACA GAGCTTGTCG AGGAGAGTTG TTGATTCATT GTTTGCCTCC CTGCTGCGGT

6370      6380      6390      6400      6410      6420
TTTTCACCGA AGTTCATGCC AGTCCAGCGT TTTTGCAGCA GAAAAGCCGC CGACTTCGGT

6430      6440      6450      6460      6470      6480
TTGCGGTGCG GAGTGAAGAT CCCTTTCTTG TTACCGCCAA CGCGCAATAT GCCTTGCGAG

6490      6500      6510      6520      6530      6540

```

DNASIS
Desmond Lark

GTCGCAAAAT CGGCGAAATT CCATACCTGT TCACCGACGA CGGCGCTGAC GCGATCAAAG

6550 6560 6570 6580 6590 6600
ACGCGGTGAT ACATATCCAG CCATGCACAC TGATACTCTT CACTCCACAT GTCGGTGTAC

6610 6620 6630 6640 6650 6660
ATTGAGTGCA GCGCGGTAA CGTATCCACG CCGTATTCGG TGATGATAAT CGGCTGATGC

6670 6680 6690 6700 6710 6720
AGTTTCTCCT GCCAGGCCAG AAGTTCTTTT TCCAGTACCT TCTCTGCCGT TTCCAAATCG

6730 6740 6750 6760 6770 6780
CCGCTTTGGA CATACCATCC GTAATAACGG TTCAGGCACA GCACATCAA GAGATCGCTG

6790 6800 6810 6820 6830 6840
ATGGTATCGG TGTGAGCGTC GCAGAACATT ACATTGACGC AGGTGATCGG ACGCGTCGGG

6850 6860 6870 6880 6890 6900
TCGAGTTTAC GCGTTGCTTC CGCCAGTGGC GCGAAATATT CCCGTGCACC TTGCGGACGG

6910 6920 6930 6940 6950 6960
GTATCCGGTT CGTTGGCAAT ACTCCACATC ACCACGCTTG GGTGGTTTTT GTCACGCGCT

6970 6980 6990 7000 7010 7020
ATCAGCTCTT TAATCGCCTG TAAGTGCCTG TGCTGAGTTT CCCCCTTGAC TGCCTCTTCG

7030 7040 7050 7060 7070 7080
CTGTACAGTT CTTTCGGCTT GTTGCCCGCT TCGAAACCAA TGCCTAAAGA GAGGTTAAAG

7090 7100 7110 7120 7130 7140
CCGACAGCAG CAGTTTCATC AATCACCACG ATGCCATGTT CATCTGCCCA GTCGAGCATC

7150 7160 7170 7180 7190 7200
TCTTCAGCGT AAGGGTAATG CGAGGTACGG TAGGAGTTGG CCCCAATCCA GTCCATTAAT

7210 7220 7230 7240 7250 7260
GCGTGGTCGT GCACCATCAG CACGTTATCG AATCCTTTGC CACGCAAGTC CGCATCTTCA

7270 7280 7290 7300 7310 7320
TGACGACCAA AGCCAGTAAA GTAGAACGGT TTGTGGTTAA TCAGGAATG TTCGCCCTTC

7330 7340 7350 7360 7370 7380
ACTGCCACTG ACCGGATGCC GACGCGAAGC GGGTAGATAT CACACTCTGT CTGGCTTTTG

7390 7400 7410 7420 7430 7440
GCTGTGACGC ACAGTTCATA GAGATAACCT TCACCCGGTT GCCAGAGGTG CGGATTCACC

7450 7460 7470 7480 7490 7500
ACTTGCAAAG TCCCGTAGT GCCTTGCCA GTTGCAACCA CCTGTTGATC CGCATCACGC

7510 7520 7530 7540 7550 7560
AGTTCAACGC TGACATCACC ATTGGCCACC ACCTGCCAGT CAACAGACGC GTGGTTACAG

7570 7580 7590 7600 7610 7620
TCTTGCGCGA CATGCGTCAC CACGGTGATA TCGTCCACCC AGGTGTTCCG CGTGGTGTAG

7630 7640 7650 7660 7670 7680
AGCATTACGC TGCGATGGAT TCCGGCATAG TTAAGAAAT CATGGAAGTA AGACTGCTTT

7690 7700 7710 7720 7730 7740
TTCTTGCCGT TTTCTGCGGT AATCACCATT CCCGGCGGGA TAGTCTGCCA GTTCAGTTCC

7750 7760 7770 7780 7790 7800
TTGTTACAC AAACGGTGAT ACCCTCGAC GGATTAAGA CTTCAAGCGG TCAACTATGA

DNASIS
Desmond Lark

```

      7810      7820      7830      7840      7850      7860
AGAAGTGTTC GTCTTCGTCC CAGTAAGCTA TGTCTCCAGA ATGTAGCCAT CCATCCTTGT

      7870      7880      7890      7900      7910      7920
CAATCAAGGC GTTGGTCGCT TCCGGATTGT TTACATAACC GGACATAATC ATAGGTCCCTC

      7930      7940      7950      7960      7970      7980
TGACACATAA TTCGCCTCTC TGATTAACGC CCAGCGTTTT CCCGGTATCC AGATCCACAA

      7990      8000      8010      8020      8030      8040
CCTTCGCTTC AAAAAATGGA ACAACTTTAC CGACCGCGCC CGGTTTATCA TCCCCCTCGG

      8050      8060      8070      8080      8090      8100
GTGTAATCAG AATAGCTGAT GTAGTCTCAG TGAGCCCATC TCCTTGTCGT ATCCCTGGAA

      8110      8120      8130      8140      8150      8160
GATGGAAGCG TTTTGCAACC GCTTCCCCGA CTTCTTTTCA AAGAGGTGCG CCCCCAGAAG

      8170      8180      8190      8200      8210      8220
ATTTCGTG TAAATTAGAT AAATCGTATT TGTCAATCAG AGTGCTTTTG GCGAAGAATG

      8230      8240      8250      8260      8270      8280
AAAATAGGGT TGGTACTAGC AACGCACTTT GAATTTTGTG ATCCTGAAGG GATCGTAAAA

      8290      8300      8310      8320      8330      8340
ACAGCTCTTC TTCAAATCTA TACATTAAGA CGACTCGAAA TCCACATATC AAATATCCGA

      8350      8360      8370      8380      8390      8400
GTGTAGTAAA CATTCCAAAA CCGTGATGGA ATGGAACAAC ACTTAAAAATC GCAGTATCCG

      8410      8420      8430      8440      8450      8460
GAATGATTTG ATTGCCAAAA ATAGGATCTC TGGCATGCGA GAATCTGACG CAGGCAGTTC

      8470      8480      8490      8500      8510      8520
TATGCGGAAG GGCCACACCC TTAGGTAACC CAGTAGATCC AGAGGAATTG TTTTGTACG

      8530      8540      8550      8560      8570      8580
CAAAGGAC TCTGGTACAA AATCGTATTC ATTAATAACCG GGAGGTAGAT GAGATGTGAC

      8590      8600      8610      8620      8630      8640
GAACGTGTAC ATCGACTGAA ATCCCTGGTA ATCCGTTTTA GAATCCATGA TAATAATTTT

      8650      8660      8670      8680      8690      8700
CTGGATTATT GGTAATTTTT TTTGCACGTT CAAAATTTTT TGCAACCECT TTTTGGAAAC

      8710      8720      8730      8740      8750      8760
AAACACTACG GTAGGCTGCG AAATGTTTAT ACTGTTGAGC AATTCACGTT CATTATAAAT

      8770      8780      8790      8800      8810      8820
GTCGTTCCGG GCGGCAACTG CAACTCCGAT AAATAACGCG CCCAACACCG GCATAAAGAA

      8830      8840      8850      8860      8870      8880
TTGAAGAGAG TTTTCACTGC ATACGACGAT TCTGTGATTT GTATTACGCC CATATCGTTT

      8890      8900      8910      8920      8930      8940
CATAGCTTCT GCCAACCGAA CGGACATTTT GAAGTATTCC GCGTACGTGA TGTTACCTTC

      8950      8960      8970      8980      8990      9000
GATATGTGCA TCTGTAAAAG GAATTGTTCC AGGAACCAGG GCGTATCTCT TCATAGCCTT

      9010      9020      9030      9040      9050      9060
ATGCAGTTGC TCTCCAGCGG TTCCATCCTC TAGCTTTGCT TCTCAATTC TTATTTGCAT

      9070      9080      9090      9100      9110      9120
AATGAGAAAA AAAGGAAAAT TAATTTTAAC ACCAATTGAG TAGTTGATTG AGCAAATGCG

```

DNASIS
Desmond Lark

```

      9130      9140      9150      9160      9170      9180
TTGCCAAAAA GGATGCTTTA GAGACAGTGT TCTCTGCACA GATAAGGACA AACATTATTC

      9190      9200      9210      9220      9230      9240
AGAGGGAGTA CCCAGAGCTG AGACTCCTAA GCCAGTGAGT GGCACAGCAT CCAGGGAGAA

      9250      9260      9270      9280      9290      9300
ATATGCTTGT CATCACCGAA GCCTGATTCC GTAGAGCCAC ACCCTGGTAA GGGCCAATCT

      9310      9320      9330      9340      9350      9360
GCTCACACAG GATAGAGAGG GCAGGAGCCA GGGCAGAGCA TATAAGGTGA GGTAGGATCA

      9370      9380      9390      9400      9410      9420
GTTGCTCCTC ACATTTGCTT CTGACATAGT TGTGTTGGGA GCTTGGATCG ATCCACCATG

      9430      9440      9450      9460      9470      9480
GGCTTCAATA CCCTGATTGA CTGGAACAGC TGTAGCCCTG AACAGCAGCG TGCCTGCTG

      9490      9500      9510      9520      9530      9540
A CGTCCGG CGATTTCCGC CTCTGACAGT ATTACCCGGA CGGTCAGCGA TATTTTGAT

      9550      9560      9570      9580      9590      9600
AATGTAAAAA CGCGCGGTGA CGATGCCCTG CGTGAATACA GCGCTAAATT TGATAAAACA

      9610      9620      9630      9640      9650      9660
GAAGTGACAG CGCTACGCGT CACCCCTGAA GAGATCGCCG CCGCCGGCGC GCGTCTGAGC

      9670      9680      9690      9700      9710      9720
GACGAATTAA AACAGGCGAT GACCGCTGCC GTCAAAAATA TTGAAACGTT CCATTCGCGC

      9730      9740      9750      9760      9770      9780
CAGACGCTAC CGCTGTAGA TGTGGAAACC CAGCCAGGCG TGCCTTGCCA GCAGGTTACG

      9790      9800      9810      9820      9830      9840
CGTCCGCTCT CGTCTGTGCG TCTGTATATT CCCGGCGGCT CGGCTCCGCT CTTCTCAACG

      9850      9860      9870      9880      9890      9900
C CTGATGC TGGCGACGCC GCGCGCATT GCGGGATGCC AGAAGGTGGT TCTGTGCTCG

      9910      9920      9930      9940      9950      9960
CCGCCGCCCA TCGCTGATGA AATCCTCTAT GCGGCGCAAC TGTGTGGCGT GCAGGAAATC

      9970      9980      9990      10000      10010      10020
TTTAACGTCT GCGGCGCGCA GCGGATTGCC GCTCTGGCCT TCGGCAGCGA GTCCGTACCG

      10030      10040      10050      10060      10070      10080
AAAGTGATA AAATTTTGG CCCCAGCAAC GCCTTTGTAA CCGAAGCCAA ACGTCAGGTC

      10090      10100      10110      10120      10130      10140
AGCCAGCGTC TCGACGGCGC GGCTATCGAT ATGCCAGCCG GGCCGTCTGA AGTACTGGTG

      10150      10160      10170      10180      10190      10200
ATCGCAGACA GCGGCGCAAC ACCGGATTTC GTCGCTTCTG ACCTGCTCTC CCAGGCTGAG

      10210      10220      10230      10240      10250      10260
CACGGCCCCG ATTCCAGGT GATCCTGTG ACGCCTGATG CTGACATTGC CCGCAAGGTG

      10270      10280      10290      10300      10310      10320
GCGGAGGCGG TAGAACGTCA ACTGGCGGAA CTGCCGCGCG CGGACACCGC CCGGCAGGCC

      10330      10340      10350      10360      10370      10380
CTGAGCGCCA GTCGTCTGAT TGTGACCAAA GATTAGCGC AGTGCCTCGC CATCTCTAAT

      10390      10400      10410      10420      10430      10440

```

DNASIS
Desmond Lark

CAGTATGGGC CGGAACACTT AATCATCCAG ACGCGCAATG CGCGCGATTT GGTGGATGCG

10450 10460 10470 10480 10490 10500
ATTACCAGCG CAGGCTCGGT ATTTCTCGGC GACTGGTCGC CGGAATCCGC CGGTGATTAC

10510 10520 10530 10540 10550 10560
GCTTCCGGAA CCAACCATGT TTTACCGACC TATGGCTATA CTGCTACCTG TTCCAGCCTT

10570 10580 10590 10600 10610 10620
GGGTTAGCGG ATTTCCAGAA ACGGATGACC GTTCAGGAAC TGTCGAAAGC GGGCTTTTCC

10630 10640 10650 10660 10670 10680
GCTCTGGCAT CAACCATTGA AACATTGGCG GCGGCAGAAC GTCTGACCGC CCATAAAAAAT

10690 10700 10710 10720 10730 10740
GCCGTGACCC TGC GCGTAAA CGCCCTCAAG GAGCAAGCAT GAGCACTGAA AACACTCTCA

10750 10760 10770 10780 10790 10800
GCGTCGCTGA CTTAGCCCGT GAAAATGTCC GCAACCTGGA GATCCAGACA TGGATAAGAT

10810 10820 10830 10840 10850 10860
ACATTGATGA GTTTGGACAA ACCACAATA GAATGCAGTG AAAAAAATGC TTTATTTGTG

10870 10880 10890 10900 10910 10920
AAATTTGTGA TGCTATTGCT TTATTTGTAA CCATTATAAG CTGCAATAAA CAAGTTAACA

10930 10940 10950 10960 10970 10980
ACAACAATTG CATTCATTTT ATGTTTCAGG TTCAGGGGGA GGTGTGGGAG GTTTTTTAAA

10990 11000 11010 11020 11030 11040
GCAAGTAAAA CCTCTACAAA TGTGGTATGG CTGATTATGA TCTCTAGGGC CGGCCCTCGA

11050 11060 11070 11080 11090 11100
CGGCGCGCCT GGCCGCTACT AACTCTCTCC TCCCTCCTTT TTCCTGCAGG CTCAAGGCGC

11110 11120 11130 11140 11150 11160
GCATGCCCCA CGGCGAGGAT CTCGTCTGTA CCCATGGCGA TGCTGTCTTG CCGAATATCA

11170 11180 11190 11200 11210 11220
TGGTGGAAAA TGGCCGCTTT TCTGGATTCA TCGACTGTGG CCGGCTGGGT GTGGCGGACC

11230 11240 11250 11260 11270 11280
GCTATCAGGA CATAGCGTTG GCTACCCGTG ATATTGCTGA AGAGCTTGGC GGCGAATGGG

11290 11300 11310 11320 11330 11340
CTGACCGCTT CCTCGTGCTT TACGGTATCG CCGCTCCCGA TTCGCAGCGC ATCGCCTTCT

11350 11360 11370 11380 11390 11400
ATCGCCTTCT TGACGAGTTC TTCTGAGCGG GACTCTGGGG TTCGAAATGA CCGACCAAGC

11410 11420 11430 11440 11450 11460
GACGCCCCAAC CTGCCATCAC GAGATTTCTGA TTCCACCGCC GCCTTCTATG AAAGGTTGGG

11470 11480 11490 11500 11510 11520
CTTCGGAATC GTTTTCCGGG ACGCCGGCTG GATGATCTC CAGCGCGGGG ATCTCATGCT

11530 11540 11550 11560 11570 11580
GGAGTTCTTC GCCCACCACA ACTTGTTTAT TGCAGCTTAT AATGGTTACA AATAAAGCAA

11590 11600 11610 11620 11630 11640
TAGCATCACA AATTTACAAA ATAAAGCATT TTTTCACTG CATTCTAGTT GTGGTTTGTG

11650 11660 11670 11680 11690 11700
CAAATCATC AATCTATCTT ATCATGTCTG GATCGCGGCC GGTCTCTCTC TAGCCCTAGG

DNASIS

18 / 51

Desmond Lark

11710 11720 11730 11740 11750 11760
TCTAGACTTG GCAGAACATA TCCATCGCGT CCGCCATCTC CAGCAGCCGC ACGCGGCGCA

11770 11780 11790 11800 11810 11820
TCTCGGGCAG CGTTGGGTCC TGGCCACGGG TGC GCATGAT CGTGCTCTG TCGTTGAGGA

11830 11840 11850 11860 11870 11880
CCCGGCTAGG CTGGCGGGGT TGCTTACTG GTTAGCAGAA TGAATCACCG ATACGCGAGC

11890 11900 11910 11920 11930 11940
GAACGTGAAG CGACTGCTGC TGCAAAACGT CTGCGACCTG AGCAACAACA TGAATGGTCT

11950 11960 11970 11980 11990 12000
TCGGTTTCCG TGTTCGTAA AGTCTGGAAA CCGGGAAGTC AGCGCCCTGC ACCATTATGT

12010 12020 12030 12040 12050 12060
TCCGGATCTG CATCGCAGGA TGCTGCTGGC TACCCTGTGG AACACCTACA TCTGTATTAA

12070 12080 12090 12100 12110 12120
CGAAGCGCTG GCATTGACCC TGAGTGATT TTCTCTGGTC CCGCCGCATC CATACCGCCA

12130 12140 12150 12160 12170 12180
GTTGTTTACC CTCACAACGT TCCAGTAACC GGGCATGTTT ATCATCAGTA ACCCGTATCG

12190 12200 12210 12220 12230 12240
TGAGCATCCT CTCTCGTTTC ATCGGTATCA TTACCCCAT GAACAGAAAT CCCCCTTACA

12250 12260 12270 12280 12290 12300
CGGAGGCATC AGTGACAAA CAGGAAAAAA CCGCCCTTAA CATGGCCCGC TTTATCAGAA

12310 12320 12330 12340 12350 12360
GCCAGACATT AACGCTTCTG GAGAACTCA ACGAGCTGGA CGCGGATGAA CAGGCAGACA

12370 12380 12390 12400 12410 12420
TCTGTGAATC GCTTCACGAC CACGCTGATG AGCTTTACCG CAGCTGCCTC GCGCGTTTCG

12430 12440 12450 12460 12470 12480
GTGATGACGG TGA AAACCTC TGACACATGC AGCTCCCGGA GACGGTCACA GCTTGTCTGT

12490 12500 12510 12520 12530 12540
AAGCGGATGC CGGGAGCAGA CAAGCCCGTC AGGGCGCGTC AGCGGGTGT TGGCGGGTGC

12550 12560 12570 12580 12590 12600
GGGGCGCAGC CATGACCCAG TCACGTAGCG ATAGCGGAGT GTATACTGGC TTA ACTATGC

12610 12620 12630 12640 12650 12660
GGCATCAGAG CAGATTGTAC TGAGAGTGCA CCATATGCGG TGTGAAATAC CGCACAGATG

12670 12680 12690 12700 12710 12720
CGTAAGGAGA AAATACCGCA TCAGGCGCTC TTCCGCTTCC TCGCTCACTG ACTCGCTGCG

12730 12740 12750 12760 12770 12780
CTCGGTGCTT CGGCTGCGGC GAGCGGTATC AGCTCACTCA AAGGCGGTAA TACGGTTATC

12790 12800 12810 12820 12830 12840
CACAGAATCA GGGGATAACG CAGGAAAGAA CATGTGAGCA AAAGGCCAGC AAAAGGCCAG

12850 12860 12870 12880 12890 12900
GAACCGTAAA AAGGCCCGCT TGCTGGCGTT TTCCATAGG CTCCGCCCCC CTGACGAGCA

12910 12920 12930 12940 12950 12960
TCACAAAAT CGACGCTCAA GTCAGAGGTG GCGAAACCCG ACAGGACTAT AAAGATACCA

12970 12980 12990 13000 13010 13020
GGCGTTTCCC CCTGGAAGCT CCCTCGTGCG CTCTCCTGTT CCGACCTGC CGTTACCGG

DNASIS
Desmond Lark

```

13030      13040      13050      13060      13070      13080
ATACCTGTCC GCCTTTCTCC CTTCCGGAAG CGTGGCGCTT TCTCATAGCT CACGCTGTAG

13090      13100      13110      13120      13130      13140
GTATCTCAGT TCGGTGTAGG TCGTTCGCTC CAAGCTGGGC TGTGTGCACG AACCCCCCGT

13150      13160      13170      13180      13190      13200
TCAGCCCGAC CGCTGCGCCT TATCCGGTAA CTATCGTCTT GAGTCCAACC CGGTAAGACA

13210      13220      13230      13240      13250      13260
CGACTTATCG CCACTGGCAG CAGCCACTGG TAACAGGATT AGCAGAGCGA GGTATGTAGG

13270      13280      13290      13300      13310      13320
CGGTGCTACA GAGTCTTGA AGTGGTGGCC TAACTACGGC TACTAGAA GGACAGTATT

13330      13340      13350      13360      13370      13380
TGGTATCTGC GCTCTGCTGA AGCCAGTTAC CTTCCGAAAA AGAGTTGGTA GCTCTTGATC

13390      13400      13410      13420      13430      13440
JCAAAACAA ACCACCGCTG GTAGCGGTGG TTTTTTGTG TGCAAGCAGC AGATTACGGC

13450      13460      13470      13480      13490      13500
CAGAAAAAAA GGATCTCAAG AAGATCCTTT GATCTTTTCT ACGGGGTCTG ACGCTCAGTG

13510      13520      13530      13540      13550      13560
GAACGAAAAC TCACGTTAAG GGATTTTGGT CATGAGATTA TCAAAAAGGA TCTTCACCTA

13570      13580      13590      13600      13610      13620
GATCCTTTTA AATTAAAAAT GAAGTTTAA ATCAATCTAA AGTATATATG AGTAACTTG

13630      13640      13650      13660      13670      13680
GTCTGACAGT TACCAATGCT TAATCAGTGA GGCACCTATC TCAGCGATCT GTCTATTTCG

13690      13700      13710      13720      13730      13740
TTCATCCATA GTTGCTGAC TCCCCGTCGT GTAGATAACT ACGATACGGG AGGGCTTACC

13750      13760      13770      13780      13790      13800
CTGGCCCC AGTGCTGCAA TGATACCGCG AGACCCACGC TCACCGGCTC CAGATTATATC

13810      13820      13830      13840      13850      13860
AGCAATAAAC CAGCCAGCCG GAAGGGCCGA GCGCAGAAGT GGTCTGCAA CTTTATCCGC

13870      13880      13890      13900      13910      13920
CTCCATCCAG TCTATTAATT GTTGCCGGGA AGCTAGAGTA AGTAGTTCGC CAGTTAATAG

13930      13940      13950      13960      13970      13980
TTTGCACAAC GTTGTTGCCA TTGCTGCAGG CATCGTGGTG TCACGCTCGT CGTTTGGTAT

13990      14000      14010      14020      14030      14040
GGCTTCATTC AGCTCCGGTT CCCAACGATC AAGGCGAGTT ACATGATCCC CCATGTTGTG

14050      14060      14070      14080      14090      14100
CAAAAAAGCG GTTAGCTCCT TCGGTCTCTC GATCGTTGTC AGAAGTAAGT TGGCCGCACT

14110      14120      14130      14140      14150      14160
GTTATCACTC ATGGTTATGG CAGCACTGCA TAATTCTCTT ACTGTCATGC CATCCGTAAG

14170      14180      14190      14200      14210      14220
ATGCTTTTCT GTGACTGGTG AGTACTCAAC CAAGTCATTG TGAGAATAGT GTATGCGGCG

14230      14240      14250      14260      14270      14280
ACCGAGTTGC TCTTGCCCGG CGTCAACACG GGATAATACC GCGCCACATA GCAGAACTTT

14290      14300      14310      14320      14330      14340
```

DNASIS
Desmond Lark

AAAAGTGCTC ATCATTGGAA AACGTTCTTC GGGGCGAAAA CTCTCAAGGA TCTTACCGCT
14350 14360 14370 14380 14390 14400
GTTGAGATCC AGTTCGATGT AACCCACTCG TGCACCCAAC TGATCTTCAG CATCTTTTAC
14410 14420 14430 14440 14450 14460
TTTCACCAGC GTTCTGCGGT GAGCAAAAAC AGGAAGGCAA AATGCCGCAA AAAAGGGAAT
14470 14480 14490 14500 14510 14520
AAGGGCGACA CGGAAATGTT GAATACTCAT ACTCTTCCTT TTTCAATATT ATTGAAGCAT
14530 14540 14550 14560 14570 14580
TTATCAGGGT TATTGTCTCA TGAGCGGATA CATATTTGAA TGTATTTAGA AAAATAAACA
14590 14600 14610 14620 14630 14640
AATAGGGGTT CCGCGCACAT TTCCCCGAAA AGTGCCACCT GACGTCTAAG AAACCATTAT
14650 14660 14670 14680 14690 14700
TATCATGACA TTAACCTATA AAAATAGGCG TATCACGAGG CCCTTTCGTC TTCAAGAA..

FIGURE 8

2/3/98

DNASIS
Molly Lark

```

      10      20      30      40      50      60
TTAATTAAGG GCGGGAGAAT GGGCGGAAGT GGGCGGAGTT AGGGGCGGGA TGGGCGGAGT

      70      80      90     100     110     120
TAGGGGCGGG ACTATGGTTG CTGACTAATT GAGATGCATG CTTTGCATAC TTCTGCCTGC

      130     140     150     160     170     180
TGGGGAGCCT GGGGACTTTC CACACCTGGT TGCTGACTAA TTGAGATGCA TGCTTTGCAT

      190     200     210     220     230     240
ACTTCTGCCT GCTGGGGAGC CTGGGGACTT TCCACACCCT AACTGACACA CATTCCACAG

      250     260     270     280     290     300
AATTAATTCC CCTAGTTATT AATAGTAATC AATTACGGGG TCATTAGTTC ATAGCCCATA

      310     320     330     340     350     360
TATGGAGTTC CGCGTTACAT AACTTACGGT AAATGGCCCG CCTGGCTGAC CGCCCAACGA

      370     380     390     400     410     420
:CCGCCCA TTGACGTCAA TAATGACGTA TGTTCCCATTA GTAACGCCAA TAGGGACTTT

      430     440     450     460     470     480
CCATTGACGT CAATGGGTGG AGTATTTACG GTAAACTGCC CACTTGGCAG TACATCAAGT

      490     500     510     520     530     540
GTATCATATG CCAAGTACGC CCCCTATTGA CGTCAATGAC GGTAATGGC CCGCCTGGCA

      550     560     570     580     590     600
TTATGCCCAG TACATGACCT TATGGGACTT TCCTACTTGG CAGTACATCT ACGTATTAGT

      610     620     630     640     650     660
CATCGCTATT ACCATGGTGA TGGGGTTTTG GCAGTACATC AATGGGCGTG GATAGCGGTT

      670     680     690     700     710     720
TGACTCACGG GGATTTCCAA GTCTCCACCC CATTGACGTC AATGGGAGTT TGTTTTGAAG

      730     740     750     760     770     780
TGGCCGGC CAGCTTTATT TAACGTGTTT ACGTCGAGTC AATTGTACAC TAACGACAGT

      790     800     810     820     830     840
GATGAAAGAA ATACAAAAGC GCATAATATT TTGAACGACG TCGAACCTTT ATTACAAAAC

      850     860     870     880     890     900
AAAACACAAA CGAATATCGA CAAAGCTAGA TTGCTGTAC AAGATTTGGC AAGTTTGTG

      910     920     930     940     950     960
GCGTTGAGCG AAAATCCATT AGATAGTCCA GCCATCGGTT CGGAAAAACA ACCCTTGTTT

      970     980     990    1000    1010    1020
GAAACTAATC GAAACCTATT TTACAAATCT ATTGAGGATT TAATATTTAA ATTCAGATAT

      1030    1040    1050    1060    1070    1080
AAAGACGCTG AAAATCATTT GATTTTCGCT CTAACATACC ACCCTAAAGA TTATAAATTT

      1090    1100    1110    1120    1130    1140
AATGAATTAT TAAAATACAT CAGCAACTAT ATATTGATAG ACATTTCCAG TTTGTGATAT

      1150    1160    1170    1180    1190    1200
TAGTTTGTGC GTCTCATTAC AATGGCTGTT ATTTTAAACA ACAAACAAC TCTCGCAGAC

      1210    1220    1230    1240    1250    1260
AATAGTATAG AAAAGGGAGG TGAAGTGTTC TTGTTTAACG GTTCGTACAA CATTTTGGA

      1270    1280    1290    1300    1310    1320
AGTTATGTTA ATCCGGTGCT GCTAAAAAAT GGTGTAATTG AACTAGAAGA AGCTGCGTAC

```

DNASIS
Molly Lark

```
1330      1340      1350      1360      1370      1380
TATGCCGGCA ACATATTGTA CAAACCGAC GATCCCAAAT TCATTGATTA TATAAATTTA

1390      1400      1410      1420      1430      1440
ATAATTAAAG CAACACACTC CGAAGAACTA CCAGAAAATA GCACTGTTGT AAATTACAGA

1450      1460      1470      1480      1490      1500
AAAACTATGC GCAGCGGTAC TATACACCCC ATTAATAAAG ACATATATAT TTATGACAAC

1510      1520      1530      1540      1550      1560
AAAAAATTTA CTCTATACGA TAGATACATA TATGGATACG ATAATAACTA TGTTAATTTT

1570      1580      1590      1600      1610      1620
TATGAGGAGA AAAATGAAAA AGAGAAGGAA TACGAAGAAG AAGACGACAA GCGCTCTAGT

1630      1640      1650      1660      1670      1680
TTATGTGAAA ATAAAATTAT ATTGTCGCAA ATTAAGTGTG AATCATTGTA AAATGATTTT

1690      1700      1710      1720      1730      1740
AAATATTACC TCAGCGATTA TAACTACGCG TTTTCAATTA TAGATAATAC TACAAATGTT

1750      1760      1770      1780      1790      1800
CTTGTTGCGT TTGGTTTGTA TCGTTAATAA AAAACAAATT TGACATTTAT AATTGTTTTA

1810      1820      1830      1840      1850      1860
TTATTCAATA ATTACAAATA GGATTGAGAC CCTTGCAATT GCCAGCAAAC GGACAGAGCT

1870      1880      1890      1900      1910      1920
TGTCGAGGAG AGTTGTTGAT TCATTGTTTG CCTCCCTGCT GCGGTTTTTC ACCGAAGTTC

1930      1940      1950      1960      1970      1980
ATGCCAGTCC AGCGTTTTTG CAGCAGAAAA GCCGCCGACT TCGGTTTGCG GTCGCGAGTG

1990      2000      2010      2020      2030      2040
AAGATCCCTT TCTTGTACC GCCAACGCGC AATATGCCCT GCGAGGTCGC AAAATCGGCG

2050      2060      2070      2080      2090      2100
AAATTCCATA CCTGTTCAAC GACGACGGCG CTGACGGGAT CAAAGACGCG GTGATACATA

2110      2120      2130      2140      2150      2160
TCCAGCCATG CACACTGATA CTCTTCACTC CACATGTCGG TGTACATTGA GTGCAGCCCG

2170      2180      2190      2200      2210      2220
GCTAACGTAT CCACGCCGTA TTCGGTGATG ATAATCGGCT GATGCAGTTT CTCCTGCCAG

2230      2240      2250      2260      2270      2280
GCCAGAAGTT CTTTTTCCAG TACCTTCTCT GCCGTTTCCA AATCGCCGCT TTGGACATAC

2290      2300      2310      2320      2330      2340
CATCCGTAAT AACGGTTCAG GCACAGCACA TCAAAGAGAT CGCTGATGGT ATCGGTGTGA

2350      2360      2370      2380      2390      2400
GCGTCGCAGA ACATTACATT GACGCAGGTG ATCGGACGCG TCGGGTCGAG TTTACGCGTT

2410      2420      2430      2440      2450      2460
GCTTCCGCCA GTGGCGCGAA ATATTCCCGT GCACCTTGCG GACGGGTATC CGGTTTCGTT

2470      2480      2490      2500      2510      2520
GCAATACTCC ACATACCAC GCTTGGGTGG TTTTGTAC GCGCTATCAG CTCTTTAATC

2530      2540      2550      2560      2570      2580
GCCTGTAAGT GCGCTTGCTG AGTTTCCCCG TTGACTGCCT CTTGCTGTGA CAGTTCTTTT

2590      2600      2610      2620      2630      2640
```

DNASIS

Molly Lark

GGCTTGTTC CCGCTTCGAA ACCAATGCCT AAAGAGAGGT TAAAGCCGAC AGCAGCAGTT

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 2650 | 2660 | 2670 | 2680 | 2690 | 2700 |
| TCATCAATCA | CCACGATGCC | ATGTTTCTCT | GCCGAGTCGA | GCATCTCTTC | AGCGTAAGGG |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 2710 | 2720 | 2730 | 2740 | 2750 | 2760 |
| TAATGCGAGG | TACGGTAGGA | GTTGGCCCCA | ATCCAGTCCA | TTAATGCGTG | GTCGTGCACC |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 2770 | 2780 | 2790 | 2800 | 2810 | 2820 |
| ATCAGCACGT | TATCGAATCC | TTTGCCACGC | AAGTCCGCAT | CTTCATGACG | ACCAAAGCCA |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 2830 | 2840 | 2850 | 2860 | 2870 | 2880 |
| GTAAAGTAGA | ACGGTTTGTG | GTTAATCAGG | AACTGTTTCG | CCTTCACTGC | CACTGACCGG |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 2890 | 2900 | 2910 | 2920 | 2930 | 2940 |
| ATGCCGACGC | GAAGCGGGTA | GATATCACAC | TCTGTCTGGC | TTTTGGCTGT | GACGCACAGT |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 2950 | 2960 | 2970 | 2980 | 2990 | 3000 |
| TTATAGAGAT | AACCTTCACC | CGGTTGCCAG | AGGTGCGGAT | TCACCACTTG | CAAAGTCCCC |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 3010 | 3020 | 3030 | 3040 | 3050 | 3060 |
| CTAGTGCCCT | GTCCAGTTGC | AACCACCTGT | TGATCCGCAT | CACGCAGTTC | AACGCTGACA |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 3070 | 3080 | 3090 | 3100 | 3110 | 3120 |
| TCACCATTTG | CCACCACCTG | CCAGTCAACA | GACGCGTGGT | TACAGTCTTG | CGCGACATGC |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 3130 | 3140 | 3150 | 3160 | 3170 | 3180 |
| GTCACCACGG | TGATATCGTC | CACCCAGGTG | TTCGGCGTGG | TGTAGAGCAT | TACGCTGCCA |

| | | | | | |
|------------|------------|------------|------------|-----------|------------|
| 3190 | 3200 | 3210 | 3220 | 3230 | 3240 |
| TGGATTCCGG | CATAGTTAAA | GAAATCATGG | AAGTAAGACT | GCTTTTCTT | GCCGTTTTCG |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 3250 | 3260 | 3270 | 3280 | 3290 | 3300 |
| TCGGTAATCA | CCATTCCCGG | CGGGATAGTC | TGCCAGTTCA | GTTCTGTGTT | CACACAAACG |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 3310 | 3320 | 3330 | 3340 | 3350 | 3360 |
| TTGATACCCC | TCGACGGATT | AAAGACTTCA | AGCGGTCAAC | TATGAAGAAG | TGTTCTCTTT |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 3370 | 3380 | 3390 | 3400 | 3410 | 3420 |
| CGTCCCAGTA | AGCTATGTCT | CCAGAATGTA | GCCATCCATC | CTTGTCAATC | AAGGCGTTGG |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 3430 | 3440 | 3450 | 3460 | 3470 | 3480 |
| TCGCTTCCGG | ATTGTTTACA | TAACCGGACA | TAATCATAGG | TCCTCTGACA | CATAATTCCG |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 3490 | 3500 | 3510 | 3520 | 3530 | 3540 |
| CTCTCTGATT | AACGCCCAGC | GTTTTCCCGG | TATCCAGATC | CACAACCTTC | GCTTCAAAAA |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 3550 | 3560 | 3570 | 3580 | 3590 | 3600 |
| ATGGAACAAC | TTTACCGACC | GCGCCCGGTT | TATCATCCCC | CTCGGGTGTA | ATCAGAATAG |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 3610 | 3620 | 3630 | 3640 | 3650 | 3660 |
| CTGATGTAGT | CTCAGTGAGC | CCATATCCTT | GTCGTATCCC | TGGAAGATGG | AAGCGTTTTG |

| | | | | | |
|------------|------------|-----------|------------|------------|------------|
| 3670 | 3680 | 3690 | 3700 | 3710 | 3720 |
| CAACCGCTTC | CCCGACTTCT | TTGGAAGAG | GTGCGCCCCC | AGAAGCAATT | TCGTGTAAAT |

| | | | | | |
|------------|------------|------------|------------|-----------|------------|
| 3730 | 3740 | 3750 | 3760 | 3770 | 3780 |
| TAGATAAATC | GTATTTGTCA | ATCAGAGTGC | TTTTGGCGAA | GAATGAAAT | AGGGTTGGTA |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 3790 | 3800 | 3810 | 3820 | 3830 | 3840 |
| CTAGCAACGC | ACTTTGAATT | TTGTAATCCT | GAAGGGATCG | TAAAAACAGC | TCTTCTTCAA |

| | | | | | |
|------------|------------|------------|------------|------------|-----------|
| 3850 | 3860 | 3870 | 3880 | 3890 | 3900 |
| ATCTATACAT | TAAGACGACT | CGAAATCCAC | ATATCAAATA | TCCGAGTGTA | GTAACATTC |

DNASIS
Molly Lark

```

3910      3920      3930      3940      3950      3960
CAAAACCGTG ATGGAATGGA ACAACACTTA AAATCGCAGT ATCCGGAATG ATTTGATTGC

3970      3980      3990      4000      4010      4020
CAAAAATAGG ATCTCTGGCA TGCAGGAATC TGACGCAGGC AGTTCTATGC GGAAGGGCCA

4030      4040      4050      4060      4070      4080
CACCCCTAGG TAACCCAGTA GATCCAGAGG AATTGTTTTG TCACGATCAA AGGACTCTGG

4090      4100      4110      4120      4130      4140
TACAAAATCG TATTCATTAA AACCGGGAGG TAGATGAGAT GTGACGAACG TGTACATCGA

4150      4160      4170      4180      4190      4200
CTGAAATCCC TGGTAATCCG TTTTAGAATC CATGATAATA ATTTTCTGGA TTATTGGTAA

4210      4220      4230      4240      4250      4260
TTTTTTTTGC ACGTTCAAAA TTTTGTGCAA CCCCTTTTGG GAAACAAACA CTACGGTAGG

4270      4280      4290      4300      4310      4320
TCGAAATG TTCATACTGT TGAGCAATTC ACGTTCATTA TAAATGTCGT TCGCGGGCGC

4330      4340      4350      4360      4370      4380
AACTGCAACT CCGATAAATA ACGCGCCCAA CACCGGCATA AAGAATTGAA GAGAGTTTTC

4390      4400      4410      4420      4430      4440
ACTGCATACG ACGATTCTGT GATTGTATT CAGCCCATAT CGTTTCATAG CTTCTGCCAA

4450      4460      4470      4480      4490      4500
CCGAACGGAC ATTTCAAGT ATTCCGGCTA CGTGATGTTT ACCTCGATAT GTGCATCTGT

4510      4520      4530      4540      4550      4560
AAAAGGAATT GTTCCAGGAA CCAGGGCGTA TCTCTTCATA GCCTTATGCA GTTGCTCTCC

4570      4580      4590      4600      4610      4620
AGCGGTTCCA TCCTCTAGCT TTGCTTCTCA ATTTCTTATT TGCATAATGA GAAAAAAGG

4630      4640      4650      4660      4670      4680
TATTAATT TTAACACCAA TTCAGTAGTT GATTGAGCAA ATGCGTTGCC AAAAAGGATG

4690      4700      4710      4720      4730      4740
CTTTAGAGAC AGTGTCTCTT GCACAGATAA GGACAAACAT TATTCAGAGG GAGTACCCAG

4750      4760      4770      4780      4790      4800
AGCTGAGACT CCTAAGCCAG TGAGTGGCAC AGCATCCAGG GAGAAATATG CTTGTCTATC

4810      4820      4830      4840      4850      4860
CCGAAGCCTG ATTCCTGAGA GCCACACCTT GGTAAGGGCC AATCTGCTCA CACAGGATAG

4870      4880      4890      4900      4910      4920
AGAGGGCAGG AGCCAGGGCA GAGCATATAA GGTGAGGTAG GATCAGTTGC TCCTCACATT

4930      4940      4950      4960      4970      4980
TGCTTCTGAC ATAGTTGTGT TGGGAGCTTG GATCGATCCA CCATGGGCTT CAATACCCTG

4990      5000      5010      5020      5030      5040
ATTGACTGGA ACAGCTGTAG CCCTGAACAG CAGCGTGCGC TGCTGACGCG TCCGGCGATT

5050      5060      5070      5080      5090      5100
TCCGCCTCTG ACAGTATTAC CCGGACGGTC AGCGATATTC TGGATAATGT AAAAAACGCG

5110      5120      5130      5140      5150      5160
GGTGACGATG CCCTGCGTGA ATACAGCGCT AAATTGATA AAACAGAAGT GACAGCGCTA

5170      5180      5190      5200      5210      5220
CGCGTCACCC CTGAAGAGAT CGCCGCCGCC GCGCGCGCTC TGAGCGACGA ATTAACACAG

```

DNASIS
Molly Lark

```

5230      5240      5250      5260      5270      5280
GCGATGACCG CTGCCGTCAA AAATATTGAA ACGTTCCATT CCGCGCAGAC GCTACCGCCT

5290      5300      5310      5320      5330      5340
GTAGATGTGG AAACCCAGCC AGGCGTGCGT TGCCAGCAGG TTACGCGTCC CGTCTCGTCT

5350      5360      5370      5380      5390      5400
GTCGGTCTGT ATATTCCCGG CGGCTCGGCT CCGCTCTTCT CAACGGTGCT GATGCTGGCG

5410      5420      5430      5440      5450      5460
ACGCCGGCGC GCATTGCGGG ATGCCAGAAG GTGGTTCTGT GCTCGCCGCC GCCCATCGCT

5470      5480      5490      5500      5510      5520
GATGAAATCC TCTATGCGGC GCAACTGTGT GGCCTGCAGG AAATCTTTAA CGTCGGCGGC

5530      5540      5550      5560      5570      5580
GCGCAGGCGA TTGCCGCTCT GGCCTTCGGC AGCGAGTCCG TACCGAAAGT GGATAAAATT

5590      5600      5610      5620      5630      5640
.TGGCCCCG GCAACGCCTT TGTAACCGAA GCCAAACGTC AGGTCAGCCA GCGTCTCGAC

5650      5660      5670      5680      5690      5700
GGCGCGGCTA TCGATATGCC AGCCGGGCGC TCTGAAGTAC TGGTGATCGC AGACAGCGGC

5710      5720      5730      5740      5750      5760
GCAACACCGG ATTTCGTGCG TTCTGACCTG CTCTCCAGG CTGAGCACGG CCCGGATTCC

5770      5780      5790      5800      5810      5820
CAGGTGATCC TGCTGACGCC TGATGCTGAC ATTGCCCGCA AGGTGGCGGA GGCGGTAGAA

5830      5840      5850      5860      5870      5880
CGTCAACTGG CGGAACTGCC GCGCGCGGAC ACCGCCCGGC AGGCCCTGAG CGCCAGTCTG

5890      5900      5910      5920      5930      5940
CTGATTGTGA CCAAAGATTT AGCGCAGTGC GTCGCCATCT CTAATCAGTA TGGGCCGGAA

5950      5960      5970      5980      5990      6000
.ACTTAATCA TCCAGACGCG CAATGCGCGC GATTTGGTGG ATGCGATTAC CAGCGCAGGC

6010      6020      6030      6040      6050      6060
TCGGTATTTC TCGGCGACTG GTCGCCGGAA TCCGCCGGTG ATTACGCTTC CGGAACCAAC

6070      6080      6090      6100      6110      6120
CATGTTTTAC CGACCTATGG CTATACTGCT ACCTGTTCCA GCCTTGGGTT AGCGGATTTT

6130      6140      6150      6160      6170      6180
CAGAAACGGA TGACCGTTCA GGAAGTGTG AAAGCGGGCT TTTCCGCTCT GGCATCAACC

6190      6200      6210      6220      6230      6240
ATTGAAACAT TGGCGGCGGC AGAAGCTCTG ACCGCCATA AAAATGCCGT GACCCTGCGC

6250      6260      6270      6280      6290      6300
GTAAACGCCC TCAAGGAGCA AGCATGAGGC ACTGAAAACA CTCTCAGCGT CGCTGACTTA

6310      6320      6330      6340      6350      6360
GCCCCGTGAAA ATGTCCGCAA CCTGGAGATC CAGACATGAT AAGATACATT GATGAGTTTG

6370      6380      6390      6400      6410      6420
GACAAACCAC AACTAGAATG CAGTGAAAAA AATGCTTTAT TTGTGAAATT TGTGATGCTA

6430      6440      6450      6460      6470      6480
TTGCTTTATT TGTAACCATT ATAAGCTGCA ATAAACAAGT TAACAACAAC AATTGCATTC

6490      6500      6510      6520      6530      6540

```

DNASIS
Molly Lark

ATTTTATGTT TCAGGTTT CAG GGGGAGGTGT GGGAGGTTTT TTAAAGCAAG TAAAACCTCT
 6550 6560 6570 6580 6590 -6600
 ACAAATGTGG TATGGCTGAT TATGATCTCT AGGGCCGGCC CTCGACGGCG CGCCTCTAGA
 6610 6620 6630 6640 6650 6660
 GCAGTGTGGT TTTGCAAGAG GAAGCAAAAA GCCTCTCCAC CCAGGCCTGG AATGTTTCCA
 6670 6680 6690 6700 6710 6720
 CCCAATGTCT AGCAGTGTGG TTTTGCAAGA GGAAGCAAAA AGCCTCTCCA CCCAGGCCTG
 6730 6740 6750 6760 6770 6780
 GAATGTTTCC ACCCAATGTC GAGCAAACCC CGCCGAGCGT CTTGTCATTG GCGAATTCTGA
 6790 6800 6810 6820 6830 6840
 ACACGCAGAT GCAGTCGGGG CGGCGCGGTC CCAGGTCCAC TTCGCATATT AAGGTGACGC
 6850 6860 6870 6880 6890 6900
 GTGTGGCCTC GAACACCGAG CGACCCTGCA GCCAATATGG GATCGGCCAT TGAACAAGAT
 6910 6920 6930 6940 6950 6960
 GGATTGCACG CAGGTTCTCC GGCCGCTTGG GTGGAGAGGC TATTCGGCTA TGAATGGGCA
 6970 6980 6990 7000 7010 7020
 CAACAGACAA TCGGCTGCTC TGATGCCGCC GTGTTCCGGC TGTCAGCGCA GGGGCGCCCG
 7030 7040 7050 7060 7070 7080
 GTTCTTTTTG TCAAGACCGA CCTGTCCGGT GCCCTGAATG AACTGCAGGT AAGTGCGGCC
 7090 7100 7110 7120 7130 7140
 GTCGATGGCC GAGGCGGCCCT CGGCCTCTGC ATAAATAAAA AAAATTAGTC AGCCATGCAT
 7150 7160 7170 7180 7190 7200
 GGGGCGGAGA ATGGGCGGAA CTGGGCGGAG TTAGGGGCGG GATGGGCGGA GTTAGGGGCG
 7210 7220 7230 7240 7250 7260
 TACTATGGT TGCTGACTAA TTGAGATGCA TGCTTTGCAT ACTTCTGCCT GCTGGGGAGC
 7270 7280 7290 7300 7310 7320
 CTGGGGACTT TCCACACCTG GTTGCTGACT AATTGAGATG CATGCTTTGC ATACTTCTGC
 7330 7340 7350 7360 7370 7380
 CTGCTGGGGA GCCTGGGGAC TTTCCACACC CTAACAGACA CACATTCCAC AGAATTAATT
 7390 7400 7410 7420 7430 7440
 CCCCTAGTTA TTAATAGTAA TCAATTACGG GGTCATTAGT TCATAGCCCA TATATGGAGT
 7450 7460 7470 7480 7490 7500
 TCCGCGTTAC ATAACCTACG GTAAATGGCC CGCCTGGCTG ACCGCCCAAC GACCCCGGCC
 7510 7520 7530 7540 7550 7560
 CATTGACGTC AATAATGACG TATGTTCCCA TAGTAACGCC AATAGGGACT TTCCATTGAC
 7570 7580 7590 7600 7610 7620
 GTCAATGGGT GGACTATTTA CGGTAAACTG CCCACTTGGC AGTACATCAA GTGTATCATA
 7630 7640 7650 7660 7670 7680
 TGCCAAGTAC GCGCCCTATT GACGTCAATG ACGGTAAATG GCGCGCCTGG CATTATGCCC
 7690 7700 7710 7720 7730 7740
 AGTACATGAC CTTATGGGAC TTTCTACTT GGCAGTACAT CTACGTATTA GTCATCGCTA
 7750 7760 7770 7780 7790 7800
 TTACCATGGT GATGCGGTTT TGGCAGTACA TCAATGGGCG TGGATAGCGG TTTGACTCAC

DNASIS
Molly Lark

7810 7820 7830 7840 7850 7860
GGGGATTTC AAGTCTCCAC CCCATTGACG TCAATGGGAG TTTGTTTTGG CACCAAAATC
7870 7880 7890 7900 7910 7920
AACGGGACTT TCCAAAATGT CGTAACAAC TCGCCCCATT GACGCAAATG GCGGGTAGGC
7930 7940 7950 7960 7970 7980
GTGTACGGTG GGAGGTCTAT ATAAGCAGAG CTGGGTACGT GAACCGTCAG ATCGCCTGGA
7990 8000 8010 8020 8030 8040
GACGCCATCA CAGATCTCTC ACTATGGATT TTCAGGTGCA GATTATCAGC TTCCTGCTAA
8050 8060 8070 8080 8090 8100
TCAGTGCTTC AGTCATAATG TCCAGAGGAC AAATTGTTCT CTCCAGTCT CCAGCAATCC
8110 8120 8130 8140 8150 8160
TGTCTGCATC TCCAGGGGAG AAGGTCAAA TGAATTGCAG GGCCAGCTCA AGTGTAAGTT
8170 8180 8190 8200 8210 8220
ATCCACTG GTTCCAGCAG AAGCCAGGAT CCTCCCCCAA ACCCTGGATT TATGCCACAT
8230 8240 8250 8260 8270 8280
CCAACCTGGC TTCTGGAGTC CCGTTTCGCT TCAGTGGCAG TGGGTCTGGG ACTTCTTACT
8290 8300 8310 8320 8330 8340
CTCTCACAAT CAGCAGAGTG GAGGCTGAAG ATGCTGCCAC TTATTACTGC CAGCAGTGGA
8350 8360 8370 8380 8390 8400
CTAGTAACCC ACCCAGTTC GGAGGGGGGA CCAAGCTGGA AATCAAACGT ACGGTGGCTG
8410 8420 8430 8440 8450 8460
CACCATCTGT CTTCATCTTC CCGCCATCTG ATGAGCAGTT GAAATCTGGA ACTGCCTCTG
8470 8480 8490 8500 8510 8520
TTGTGTGCCT GCTGAATAAC TTCTATCCCA GAGAGGCCAA AGTACAGTGG AAGGTGGATA
8530 8540 8550 8560 8570 8580
TCCCCTCCA ATCGGGTAAC TCCCAGGAGA GTGTACAGA GCAGGACAGC AAGGACAGCA
8590 8600 8610 8620 8630 8640
CCTACAGCCT CAGCAGCACC CTGACGCTGA GCAAAGCAGA CTACGAGAAA CACAAAGTCT
8650 8660 8670 8680 8690 8700
ACGCCTGCGA AGTCACCCAT CAGGGCCTGA GCTCGCCCGT CACAAAGAGC TTCAACAGGG
8710 8720 8730 8740 8750 8760
GAGAGTGTG AATTCAATC CGTTAACGGT TACCAACTAC CTAGACTGGA TTCGTGACAA
8770 8780 8790 8800 8810 8820
CATGGGCGG TGATATCTAC GTATGATCAG CCTCGACTGT GCCTTCTAGT TGCCAGCCAT
8830 8840 8850 8860 8870 8880
CTGTTGTTTT CCCCTCCCC GTGCCTTCTT TGACCCTGGA AGGTGCCACT CCCACTGTCC
8890 8900 8910 8920 8930 8940
TTTCCTAATA AAATGAGGAA ATTGCATCGC ATTGTCTGAG TAGGTGTCAT TCTATTCTGG
8950 8960 8970 8980 8990 9000
GGGGTGGGGT GGGGCAGGAC AGCAAGGGGG AGGATTGGGA AGACAATAGC AGGCATGCTG
9010 9020 9030 9040 9050 9060
GGGATGCGGT GGGCTCTATG GAACCAGCTG GGGCTCGACA GCTATGCCAA GTACGCCCCC
9070 9080 9090 9100 9110 9120
TATTGACGTC AATGACGGTA AATGGCCCCG CTGGCATTAT GCCCAGTACA TGACCTTATG

DNASIS
Molly Lark

```

      9130      9140      9150      9160      9170      9180
GGACTTTCCT ACTTGGCAGT ACATCTACGT ATTAGTCATC GCTATTACCA TGGTGATGCG

      9190      9200      9210      9220      9230      9240
GTTTTGGCAG TACATCAATG GGCCTGGATA GCGGTTTGAC TCACGGGGAT TTCCAAGTCT

      9250      9260      9270      9280      9290      9300
CCACCCCATG GACGTCAATG GGAGTTTGTG TTGGCACCAA AATCAACGGG ACTTTCCAAA

      9310      9320      9330      9340      9350      9360
ATGTCGTAAC AACTCCGCCC CATTGACGCA AATGGGCGGT AGGCGTGTAC GGTGGGAGGT

      9370      9380      9390      9400      9410      9420
CTATATAAGC AGAGCTGGGT ACGTCCTCAC ATTCAGTGAT CAGCACTGAA CACAGACCCG

      9430      9440      9450      9460      9470      9480
TCGACATGGG TTGGAGCCTC ATCTTGCTCT TCCTTGTCGC TGTGCTACG CGTGTCTGT

      9490      9500      9510      9520      9530      9540
CCCAGGTACA ACTGCAGCAG CCTGGGGCTG AGCTGGTGAA GCCTGGGGCC TCAGTGAAGA

      9550      9560      9570      9580      9590      9600
TGTCCTGCAA GGCTTCTGGC TACACATTTA CCAGTTACAA TATGCACTGG GTAAACAGA

      9610      9620      9630      9640      9650      9660
CACCTGGTCG GGGCCTGGAA TGGATTGGAG CTATTTATCC CGGAAATGGT GATACTTCCT

      9670      9680      9690      9700      9710      9720
ACAATCAGAA GTTCAAAGGC AAGGCCACAT TGA CTGCAGA CAAATCCTCC AGCACAGCCT

      9730      9740      9750      9760      9770      9780
ACATGCAGCT CAGCAGCCTG ACATCTGAGG ACTCTGCGGT CTATTACTGT GCAAGATCGA

      9790      9800      9810      9820      9830      9840
CTTACTACGG CCGTGACTGG TACTTCAATG TCTGGGGCGC AGGGACCACG GTCACCGTCT

      9850      9860      9870      9880      9890      9900
CTGCAGCTAG CACCAAGGGC CCATCGGTCT TCCCCCTGGC ACCCTCCTCC AAGAGCACCT

      9910      9920      9930      9940      9950      9960
CTGGGGGGCAC AGCGGCCCTG GGCTGCCTGG TCAAGGACTA CTCCCCGAA CCGGTGACGG

      9970      9980      9990      10000      10010      10020
TGTCGTGGAA CTCAGGCGCC CTGACCAGCG GCGTGACAC CTTCCTGGCT GTCCTACAGT

      10030      10040      10050      10060      10070      10080
CCTCAGGACT CTA CTCCCTC AGCAGCGTGG TGACCGTGCC CTCCAGCAGC TTGGGGACCC

      10090      10100      10110      10120      10130      10140
AGACCTACAT CTGCAACGTG AATCACAAGC CCAGCAACAC CAAGGTGGAC AAGAAAGCAG

      10150      10160      10170      10180      10190      10200
AGCCCAAATC TTGTGACAAA ACTCACACAT GCCCACCCTG CCCAGCACCT GAACTCCTGG

      10210      10220      10230      10240      10250      10260
GGGGACCCTC AGTCTTCTC TTCCCCCAA AACCCAAGGA CACCCTCATG ATCTCCCGGA

      10270      10280      10290      10300      10310      10320
CCCCTGAGGT CACATGCGTG GTGGTGGACG TGAGCCACGA AGACCCTGAG GTCAAGTTCA

      10330      10340      10350      10360      10370      10380
ACTGGTACGT GGACGGCGTG GAGGTGCATA ATGCCAAGAC AAAGCCGCGG GAGGAGCAGT

      10390      10400      10410      10420      10430      10440

```

DNASIS
Molly Lark

ACAAACAGCAC GTACCGTGTG GTCAGCGTCC TCACCGTCCT GCACCAGGAC TGGCTGAATG
10450 10460 10470 10480 10490 10500
GCAAGGAGTA CAAGTGCAAG GTCTCCAACA AAGCCCTCCC AGCCCCATC GAGAAAACCA
10510 10520 10530 10540 10550 10560
TCTCCAAAGC CAAAGGGCAG CCCCAGAAC CACAGGTGTA CACCCTGCCC CCATCCCCGG
10570 10580 10590 10600 10610 10620
ATGAGCTGAC CAAGAACCAG GTCAGCCTGA CCTGCCTGGT CAAAGGCTTC TATCCACGG
10630 10640 10650 10660 10670 10680
ACATCGCCGT GGAGTGGGAG AGCAATGGGC AGCCGGAGAA CAACTACAAG ACCACGCCTC
10690 10700 10710 10720 10730 10740
CCGTGCTGGA CTCCGACGGC TCCTTCTTCC TCTACAGCAA GCTCACCGTG GACAAGAGCA
10750 10760 10770 10780 10790 10800
CGTGGCAGCA GGGGAACGTC TTCTCATGCT CCGTGATGCA TGAGGCTCTG CACAACCACT
10810 10820 10830 10840 10850 10860
ACACGCAGAA GAGCCTCTCC CTGTCTCCGG GTAAATGAGG ATCCGTTAAC GGTTACCAAC
10870 10880 10890 10900 10910 10920
TACCTAGACT GGATTCGTGA CAACATGCGG CCGTGATATC TACGTATGAT CAGCCTCGAC
10930 10940 10950 10960 10970 10980
TGTGCCCTTCT AGTTGCCAGC CATCTGTTGT TTGCCCTCC CCCGTGCCTT CCTTGACCCT
10990 11000 11010 11020 11030 11040
GGAAGGTGCC ACTCCCACTG TCCTTTCTTA ATAAATGAG GAAATTGCAT CGCATTGTCT
11050 11060 11070 11080 11090 11100
GAGTAGGTGT CATTCTATTC TGGGGGGTGG GGTGGGGCAG GACAGCAAGG GGGAGGATTG
11110 11120 11130 11140 11150 11160
GAAGACAAT AGCAGGCATG CTGGGGATGC GGTGGGCTCT ATGGAACCAG CTGGGGCTCG
11170 11180 11190 11200 11210 11220
ACAGCAACGC TAGGTCGAGG CCGCTACTAA CTCTCTCCTC CCTCCTTTT CCTGCAGGAC
11230 11240 11250 11260 11270 11280
GAGGCAGCGC GGCTATCGTG GCTGGCCACG ACGGGCGTTC CTTGCGCAGC TGTGCTCGAC
11290 11300 11310 11320 11330 11340
GTTGTCACTG AAGCGGGAAG GGA CTGGCTG CTATTGGGCG AAGTGCCGGG GCAGGATCTC
11350 11360 11370 11380 11390 11400
CTGTCACTC ACCTTGCTCC TGCCGAGAAA GTATCCATCA TGGCTGATGC AATGCGGCGG
11410 11420 11430 11440 11450 11460
CTGCATAGC TTGATCCGGC TACCTGCCCA TTCGACCACC AAGCGAAACA TCGCATCGAG
11470 11480 11490 11500 11510 11520
CGAGCACGTA CTCGGATGGA AGCCGGTCTT GTCGATCAGG ATGATCTGGA CGAAGAGCAT
11530 11540 11550 11560 11570 11580
CAGGGGCTCG CGCCAGCCGA ACTGTTCCGC AGGTAAGTGA GCTCCAATTC AAGCTTCTTA
11590 11600 11610 11620 11630 11640
GGGCGGCCAG CTAGTAGCTT TGCTTCTCAA TTTCTTATT GCATAATGAG AAAAAAGGA
11650 11660 11670 11680 11690 11700
AAATTAATTT TAACACCAAT TCAGTAGTTG ATTGAGCAAA TGC GTTGCCA AAAAGGATGC

DNASIS
Molly Lark

```

11710      11720      11730      11740      11750      11760
TTTAGAGACA GTGTTCTCTG CACAGATAAG GACAAACATT ATTCAGAGGG AGTACCCAGA

11770      11780      11790      11800      11810      11820
GCTGAGACTC CTAAGCCAGT GAGTGGCACA GCATCCAGGG AGAAATATGC TTGTCATCAC

11830      11840      11850      11860      11870      11880
CGAAGCCTGA TTCCGTAGAG CCACACCCTG GTAAGGGCCA ATCTGCTCAC ACAGGATAGA

11890      11900      11910      11920      11930      11940
GAGGGCAGGA GCCAGGGCAG AGCATATAAG GTGAGGTAGG ATCAGTTGCT CCTCACATTT

11950      11960      11970      11980      11990      12000
GCTTCTGACA TAGTTGTGTT GGGAGCTTGG ATAGCTTGGG GGGGGGACAG CTCAGGGCTG

12010      12020      12030      12040      12050      12060
CGATTTGCGC CCAAACCTGA CGGCAATCCT AGCGTGAAGG CTGGTAGGAT TTTATCCCCG

12070      12080      12090      12100      12110      12120
GCCATCAT GGTTCGACCA TTGAACGTCA TCGTCGCCGT GTCCCAAAAT ATGGGGATTG

12130      12140      12150      12160      12170      12180
GCAAGAACGG AGACCTACCC TGGCCTCCGC TCAGGAACGA GTTCAAGTAC TTCCAAAGAA

12190      12200      12210      12220      12230      12240
TGACCACAAC CTCTTCAGTG GAAGGTAAAC AGAATCTGGT GATTATGGGT AGGAAAACCT

12250      12260      12270      12280      12290      12300
GGTCTCCAT TCCTGAGAAG AATCGACCTT TAAAGGACAG AATTAATATA GTTCTCAGTA

12310      12320      12330      12340      12350      12360
GAGAACTCAA AGAACCACCA CGAGGAGCTC ATTTTCTTGC CAAAAGTTTG GATGATGCCT

12370      12380      12390      12400      12410      12420
TAAGACTTAT TGAACAACCG GAATTGGCAA GTAAAGTAGA CATGGTTTGG ATAGTCGGAG

12430      12440      12450      12460      12470      12480
AGTTCTGT TTACCAGGAA GCCATGAATC AACCAGGCCA CCTCAGACTC TTTGTGACAA

12490      12500      12510      12520      12530      12540
GGATCATGCA GGAATTTGAA AGTGACACGT TTTTCCAGA AATTGATTTG GGGAAATATA

12550      12560      12570      12580      12590      12600
AACTTCTCCC AGAATACCCA GGCCTCCTCT CTGAGGTCCA GGAGGAAAAA GGCATCAAGT

12610      12620      12630      12640      12650      12660
ATAAGTTTGA AGTCTACGAG AAGAAAGACT AACAGGAAGA TGCTTTCAAG TTCTCTGCTC

12670      12680      12690      12700      12710      12720
CCCTCCTAAA GCTATGCATT TTTATAAGAC CATGGGACTT TTGCTGGCTT TAGATCAGCC

12730      12740      12750      12760      12770      12780
TCGACTGTGC CTTCTAGTTG CCAGCCATCT GTTGTGTTGCC CCTCCCCCGT GCCTTCCTTG

12790      12800      12810      12820      12830      12840
ACCCGTGGAAG GTGCCACTCC CACTGTCCTT TCCTAATAAA ATGAGGAAAT TGCATCGCAT

12850      12860      12870      12880      12890      12900
TGTCTGAGTA GGTGTCATTG TATTCTGGGG GGTGGGGTGG GGCAGGACAG CAAGGGGGAG

12910      12920      12930      12940      12950      12960
GATTGGGAAG ACAATAGCAG GCATGCTGGG GATGCGGTGG GCTCTATGGC TTCTGAGGCG

12970      12980      12990      13000      13010      13020
GAAAGAACCA GCTGGGGCTC GAAGCGGCCG CCCATTTTCG TGGTGGTCAG ATGCGGGGATG

```

DNASIS
Molly Lark

```

13030      13040      13050      13060      13070      13080
GCGTGGGACG CGGCGGGGAG CGTCACACTG AGGTTTTCCG CCAGACGCCA CTGCTGCCAG

13090      13100      13110      13120      13130      13140
GCGCTGATGT GCCCGGCTTC TGACCATGCG GTCGCGTTCG GTTGCACTAC GCGTACTGTG

13150      13160      13170      13180      13190      13200
AGCCAGAGTT GCCCGGCGCT CTCCGGCTGC GGTAGTTCAG GCAGTTCAT CAACTGTTTA

13210      13220      13230      13240      13250      13260
CCTTGTGGAG CGACATCCAG AGGCACTTCA CCGCTTGCCA GCGGCTTACC ATCCAGCGCC

13270      13280      13290      13300      13310      13320
ACCATCCAGT GCAGGAGCTC GTTATCGCTA TGACGGAACA GGTATTCGCT GGTCACTTCG

13330      13340      13350      13360      13370      13380
ATGCTTTGCC CGGATAAACG GAACTGGAAA AACTGCTGCT GGTGTTTTGC TTCCGTCAGC

13390      13400      13410      13420      13430      13440
C .GGATGCG GCGTGCGGTC GGCAAAGACC AGACCGTTCA TACAGAACTG GCGATCGTTC

13450      13460      13470      13480      13490      13500
GGCGTATCGC CAAAATCACC GCCGTAAGCC GACCACGGGT TGCCGTTTTC ATCATATTTA

13510      13520      13530      13540      13550      13560
ATCAGCGACT GATCCACCCA GTCCAGACG AAGCCGCCCT GTAAACGGGG ATACTGACGA

13570      13580      13590      13600      13610      13620
AACGCCTGCC AGTATTTAGC GAAACCGCCA AGACTGTTAC CCATCGCGTG GCGGTATTGC

13630      13640      13650      13660      13670      13680
CAAAGGATCA GCGGGCGCGT CTCTCCAGGT AGCGAAAGCC ATTTTTGAT GGACCATTTC

13690      13700      13710      13720      13730      13740
GGCAGACCCG GGAAGGGCTG GTCTTCATCC ACGCGCGCGT ACATCGGGCA AATAATATCG

13750      13760      13770      13780      13790      13800
C .GGCCGTGG TGTCGGCTCC GCCGCCTTCA TACTGCACCG GCGGGAAGG ATCGACAGAT

13810      13820      13830      13840      13850      13860
TTGATCCAGC GATACAGCGC GTCGTGATTA GCGCCGTGGC CTGATTCAAT CCCAGCGGAC

13870      13880      13890      13900      13910      13920
CAGATGATCA CACTCGGGTG ATTACGATCG CGCTGCACCA TTCGCGTTAC GCGTTCGCTC

13930      13940      13950      13960      13970      13980
ATCGCCGGTA GCCAGCGCGG ATCATCGGTC AGACGATTCA TTGGCACCAT GCGGTGGGTT

13990      14000      14010      14020      14030      14040
TCAATATTGG CTTTCATCCAC CACATACAGG CCGTAGCGGT CGCACAGCGT GTACCACAGC

14050      14060      14070      14080      14090      14100
GGATGGTTCC GATAATGCGA ACAGCGCACG GCGTTAAAGT TGTTCTGCTT CATCAGCAGG

14110      14120      14130      14140      14150      14160
ATATCTGCA CCATCGTCTG CTCATCCATG ACCTGACCAT GCAGAGGATG ATGCTCGTGA

14170      14180      14190      14200      14210      14220
CGGTAAACGC CTCGAATCAG CAACGGCTTG CCGTTCAGCA GCAGCAGACC ATTTTCAATC

14230      14240      14250      14260      14270      14280
CGCACCTCGC GGAAACCGAC ATCGCAGGCT TCTGCTTCAA TCAGCGTGCC GTCGGCGGTG

14290      14300      14310      14320      14330      14340

```

DNASIS
Molly Lark

TGCAGTTCAA CCACCGCAG ATAGAGATTC GGGATTTCGG CGCTCCACAG TTTCGGGTTT
14350 14360 14370 14380 14390 14400
TCGACGTTCA GACGTAGTGT GACGCGATCG GCATAACCAC CACGCTCATC GATAATTTCA
14410 14420 14430 14440 14450 14460
CCGCCGAAAG GCGCGGTGCC GCTGGCGACC TCGTTTCAC CCTGCCATAA AGAAACTGTT
14470 14480 14490 14500 14510 14520
ACCCGTAGGT AGTCACGCAA CTCGCCGCAC ATCTGAACTT CAGCCTCCAG TACAGCGCGG
14530 14540 14550 14560 14570 14580
CTGAAATCAT CATTAAAGCG AGTGGCAACA TGGAAATCGC TGATTTGTGT AGTCGGTTTA
14590 14600 14610 14620 14630 14640
TGCAGCAACG AGACGTCACG GAAATGCGG CTCATCCGCC ACATATCCTG ATCTTCAGA
14650 14660 14670 14680 14690 14700
TAACTGCCGT CACTCCAGCG CAGCACCATC ACCGCGAGGC GGTTTTCTCC GGC6GTA
14710 14720 14730 14740 14750 14760
AATGCGCTCA GGTCAAATTC AGACGGCAAA CGACTGTCTT GGCCGTAACC GACCCAGCGC
14770 14780 14790 14800 14810 14820
CCGTGCGACC ACAGATGAAA CGCCGAGTTA ACGCCATCAA AAATAATTCG CGTCTGGCCT
14830 14840 14850 14860 14870 14880
TCCTGTAGCC AGCTTTCATC AACATTAAAT GTGAGCGAGT AACAAACCCGT CGGATTCTCC
14890 14900 14910 14920 14930 14940
GTGGGAACAA ACGGCGGATT GACCGTAATG GGATAGGTGA CGTTGGTGTG GATGGGCGCA
14950 14960 14970 14980 14990 15000
TCGTAACCGT GCATCTGCCA GTTGAGGGG ACGACGACAG TATCGGCTTC AGGAAGATCG
15010 15020 15030 15040 15050 15060
CACTCCAGCC AGCTTTCGGG CACCGTTTCT GGTGCCGGAA ACCAGGCAAA GCGCCATTCC
15070 15080 15090 15100 15110 15120
CCATTACGGC TCGGCAACTG TTGGGAAGGG CGATCGGTGC GGGCCTCTTC GCTATTACGC
15130 15140 15150 15160 15170 15180
CAGCTGGCGA AAGGGGGATG TGCTGCAAGG CGATTAAATT GGGTAACGCC AGGGTTTTCC
15190 15200 15210 15220 15230 15240
CAGTCACGAC GTTGTAATAAC GACTTAATCC GTCGAGGGGC TGCTCGAAG CAGACGACCT
15250 15260 15270 15280 15290 15300
TCCGTTGTGC AGCCAGCGGC GCCTGCGCCG GTGCCCAAA TCGTGCGCGA ACAAACATAA
15310 15320 15330 15340 15350 15360
CCAGAACAAA TTATACCGGC GGCACCGCCG CCACCACCTT CTCCCGTGCC TAACATTCCA
15370 15380 15390 15400 15410 15420
GC6CCTCCAC CACCACCACC ACCATCGATG TCTGAATTGC CGCCCGCTCC ACCAATGCCG
15430 15440 15450 15460 15470 15480
ACGGAACCTC AACCCGCTGC ACCTTTAGAC GACAGACAAC AATTGTTGGA AGCTATTAGA
15490 15500 15510 15520 15530 15540
AACGAAAAAA ATCGCACTCG TCTCAGACCG GTCAAACCAA AAACGGCGCC CGAAACCAGT
15550 15560 15570 15580 15590 15600
ACAATAGTTG AGGTGCCGAC TGTGTTGCCT AAAGAGACAT TTGAGCCTAA ACCGCCGTCT

DNASIS
Molly Lark

15610 15620 15630 15640 15650 15660
GCATCACC GC CACCACCTCC GCCTCCGCCT CCGCCGCCAG CCCCGCCTGC GCCTCCACCG

15670 15680 15690 15700 15710 15720
ATGGTAGATT TATCATCAGC TCCACCACCG CCGCCATTAG TAGATTGTCC GTCTGAAATG

15730 15740 15750 15760 15770 15780
TTACCACCGC CTGCACCATC GCTTTCTAAC GTGTTGTCTG AATTAAAATC GGGCACAGTT

15790 15800 15810 15820 15830 15840
AGATTGAAAC CCGCCCAAAA ACGCCCGCAA TCAGAAATAA TTCCAAAAAG CTCAACTACA

15850 15860 15870 15880 15890 15900
AATTTGATCG CGGACGTGTT AGCCGACACA ATTAATAGGC GTCGTGTGGC TATGGCAAAA

15910 15920 15930 15940 15950 15960
TCGTCTTCGG AAGCAACTTC TAACGACGAG GGTGCGGACG ACGACGATAA TCGGCCTAAT

15970 15980 15990 16000 16010 16020
AGCTAACA CGCCCGATGT TAAATATGTC CAAGCTACTA GTGGTACCGC TTGGCAGAAC

16030 16040 16050 16060 16070 16080
ATATCCATCG CGTCCGCCAT CTCCAGCAGC CGCACGCGGC GCATCTCGGG CAGCGTTGGG

16090 16100 16110 16120 16130 16140
TCCTGGCCAC GGGTGCGCAT GATCGTGCTC CTGTCGTTGA GGACCCGGCT AGGCTGGCGG

16150 16160 16170 16180 16190 16200
GGTTGCCTTA CTGTTAGCA GAATGAATCA CCGATACGCG AGCGAAGCTG AAGCGACTGC

16210 16220 16230 16240 16250 16260
TGCTGCAAAA CGTCTGCGAC CTGAGCAACA ACATGAATGG TCTTCGGTTT CCGTGTTTCG

16270 16280 16290 16300 16310 16320
TAAAGTCTGG AAACGCGGAA GTCAGCGCCC TGCACCATTA TGTTCGGAT CTGCATCGCA

16330 16340 16350 16360 16370 16380
GATGCTGCT GGCTACCCTG TGAACACCT ACATCTGTAT TAACGAAGCG CTGGCATTGA

16390 16400 16410 16420 16430 16440
CCCTGAGTGA TTTTCTCTG GTCCCGCCGC ATCCATACCG CCAGTTGTTT ACCCTCACAA

16450 16460 16470 16480 16490 16500
CGTTCCAGTA ACCGGGCATG TTCATCATCA GTAACCCGTA TCGTGAGCAT CCTCTCTCGT

16510 16520 16530 16540 16550 16560
TTCATCGGTA TCATTACCCC CATGAACAGA AATCCCCCTT ACACGGAGGC ATCAGTGACC

16570 16580 16590 16600 16610 16620
AAACAGGAAA AAACCGCCCT TAACATGGCC CGCTTTATCA GAAGCCAGAC ATTAACGCTT

16630 16640 16650 16660 16670 16680
CTGGAGAAAC TCAACGAGCT GGACGCGGAT GAACAGGCAG ACATCTGTGA ATCGCTTCAC

16690 16700 16710 16720 16730 16740
GACCACGCTG ATGAGCTTTA CCGCAGCTGC CTCGCGCGTT TCGGTGATGA CCGTGAAAAC

16750 16760 16770 16780 16790 16800
CTCTGACACA TGCAGCTCCC GGAGACGGTC ACAGCTTGTC TGTAAGCGGA TGCCGGGAGC

16810 16820 16830 16840 16850 16860
AGACAAGCCC GTCAGGGCGC GTCAGCGGGT GTTGGCGGGT GTCGGGGCGC AGCCATGACC

16870 16880 16890 16900 16910 16920
CAGTCACGTA GCGATAGCGG AGTGATACT GGCTTAACTA TGCGGCATCA GAGCAGATTG

DNASIS
Molly Lark

| | | | | | |
|------------|------------|------------|------------|-------------|-------------|
| 16930 | 16940 | 16950 | 16960 | 16970 | 16980 |
| TACTGAGAGT | GCACCATATG | CGGTGTGAAA | TACCGCACAG | ATGCGTAAGG | AGAAAATACC |
| 16990 | 17000 | 17010 | 17020 | 17030 | 17040 |
| GCATCAGGCG | CTCTCCGCT | TCCTCGCTCA | CTGACTCGCT | GCGCTCGGTC | GTTCCGGCTGC |
| 17050 | 17060 | 17070 | 17080 | 17090 | 17100 |
| GGCGAGCGGT | ATCAGCTCAC | TCAAAGGCGG | TAATACGGTT | ATCCACAGAA | TCAGGGGATA |
| 17110 | 17120 | 17130 | 17140 | 17150 | 17160 |
| ACGCAGGAAA | GAACATGTGA | GCAAAAGGCC | AGCAAAAGGC | CAGGAACCGT | AAAAAGGCCG |
| 17170 | 17180 | 17190 | 17200 | 17210 | 17220 |
| CGTTGCTGGC | GTTTTTCCAT | AGGCTCCGCC | CCCCTGACGA | GCATCACAAA | AATCGACGCT |
| 17230 | 17240 | 17250 | 17260 | 17270 | 17280 |
| CAAGTCAGAG | GTGGCGAAAC | CCGACAGGAC | TATAAAGATA | CCAGGCGTTT | CCCCCTGAA |
| 17290 | 17300 | 17310 | 17320 | 17330 | 17340 |
| GCTCCCTCGT | GCGCTCTCT | GTTCCGACCC | TGCCGCTTAC | CGGATACCTG | TCCGCCTTTC |
| 17350 | 17360 | 17370 | 17380 | 17390 | 17400 |
| TCCCTTCGGG | AAGCGTGGCG | CTTCTCATA | GCTCACGCTG | TAGGTATCTC | AGTTCGGTGT |
| 17410 | 17420 | 17430 | 17440 | 17450 | 17460 |
| AGGTGCTTCG | CTCCAAGCTG | GGCTGTGTGC | ACGAACCCCC | CGTTCAGCCC | GACCGCTGCG |
| 17470 | 17480 | 17490 | 17500 | 17510 | 17520 |
| CCTTATCCGG | TAACTATCGT | CTTGAGTCCA | ACCCGGTAAG | ACACGACTTA | TCGCCACTGG |
| 17530 | 17540 | 17550 | 17560 | 17570 | 17580 |
| CAGCAGCCAC | TGGTAACAGG | ATTAGCAGAG | CGAGGTATGT | AGGCGGTGCT | ACAGAGTTCT |
| 17590 | 17600 | 17610 | 17620 | 17630 | 17640 |
| TGAAGTGGTG | GCCTAACTAC | GGCTACACTA | GAAGGACAGT | ATTTGGTATC | TGCGCTCTGC |
| 17650 | 17660 | 17670 | 17680 | 17690 | 17700 |
| TGAAGCCAGT | TACCTTCGGA | AAAAGAGTTG | GTAGCTCTTG | ATCCGGGAAA | CAAACCACCG |
| 17710 | 17720 | 17730 | 17740 | 17750 | 17760 |
| CTGGTAGCGG | TGGTTTTTTT | GTTTGCAAGC | AGCAGATTAC | GCGCAGAAAA | AAAGGATCTC |
| 17770 | 17780 | 17790 | 17800 | 17810 | 17820 |
| AAGAAGATCC | TTTGATCTTT | TCTACGGGGT | CTGACGCTCA | GTGGAACGAA | AACTCACGTT |
| 17830 | 17840 | 17850 | 17860 | 17870 | 17880 |
| AAGGGATTTT | GGTCATGAGA | TTATCAAAAA | GGATCTTCAC | CTAGATCCTT | TAAATTAAA |
| 17890 | 17900 | 17910 | 17920 | 17930 | 17940 |
| AATGAAGTTT | TAAATCAATC | TAAAGTATAT | ATGAGTAAAC | TTGGTCTGAC | AGTTACCAAT |
| 17950 | 17960 | 17970 | 17980 | 17990 | 18000 |
| GCTTAATCAG | TGAGGCACCT | ATCTCAGCGA | TCTGTCTATT | TCGTTTCATCC | ATAGTTGCCT |
| 18010 | 18020 | 18030 | 18040 | 18050 | 18060 |
| GACTCCCCGT | CGTGTAGATA | ACTACGATAC | GGGAGGGCTT | ACCATCTGGC | CCCAGTGCTG |
| 18070 | 18080 | 18090 | 18100 | 18110 | 18120 |
| CAATGATACC | GCGAGACCCA | CGCTCACCGG | CTCCAGATTT | ATCAGCAATA | AACCAGCCAG |
| 18130 | 18140 | 18150 | 18160 | 18170 | 18180 |
| CCGGAAGGGC | CGAGCGCAGA | AGTGGTCCTG | CAACTTTATC | GCGCTCCATC | CAGTCTATTA |
| 18190 | 18200 | 18210 | 18220 | 18230 | 18240 |

DNASIS

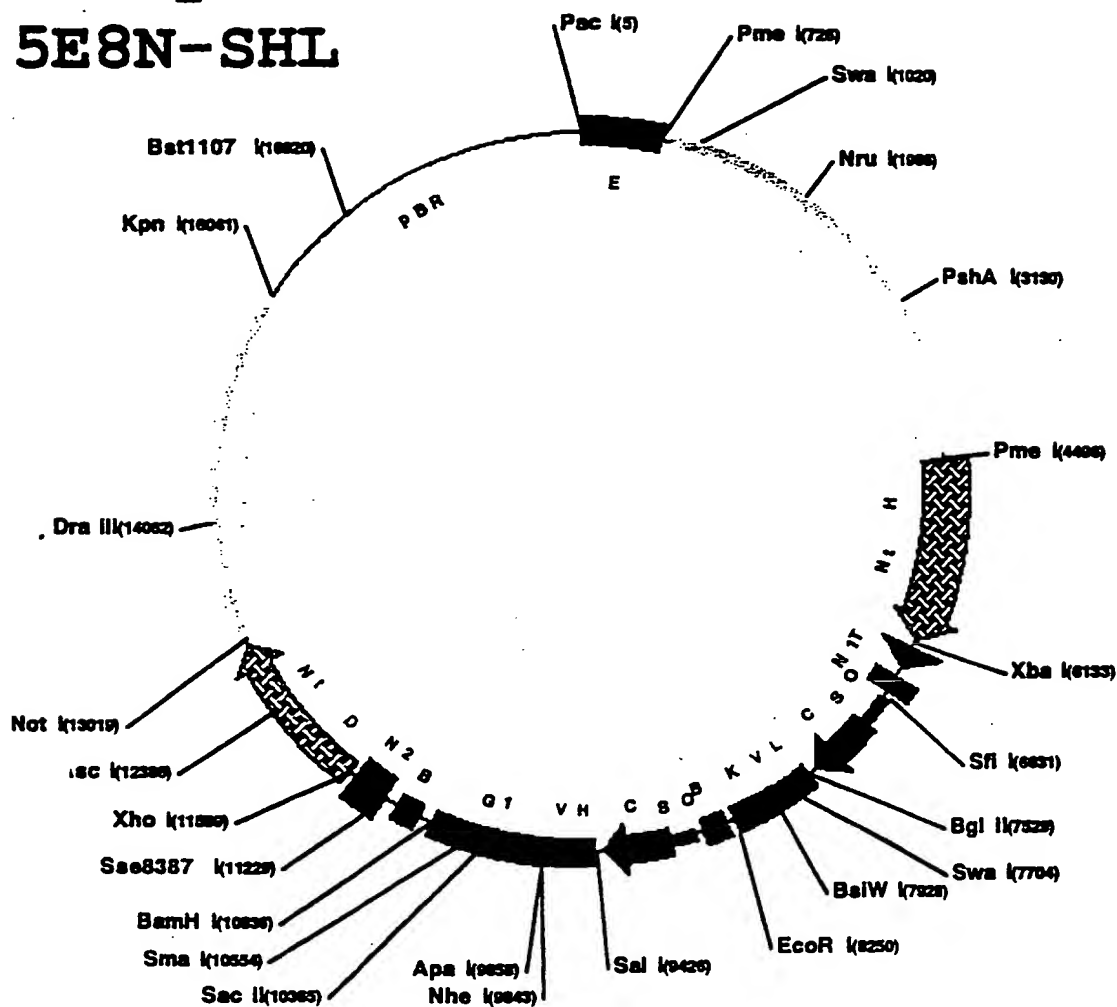
Molly Lark

ATTGTTGCCG GGAAGCTAGA GTAAGTAGTT CGCCAGTTAA TAGTTTGCGC AACGTTGTTG

18250 18260 18270 18280 18290 18300
CCATTGCTGC AGGCATCGTG GTGTCACGCT CGTCGTTTGG TATGGCTTCA TTCAGCTCCG18310 18320 18330 18340 18350 18360
GTTCCCAACG ATCAAGGCGA GTTACATGAT CCCCCATGTT GTGCAAAAAA GCGGTTAGCT18370 18380 18390 18400 18410 18420
CCTTCGGTCC TCCGATCGTT GTCAGAAAGTA AGTTGGCCGC AGTGTATCA CTCATGGTTA18430 18440 18450 18460 18470 18480
TGGCAGCACT GCATAATTCT CTTACTGTCA TGCCATCCGT AAGATGCTTT TCTGTGACTG18490 18500 18510 18520 18530 18540
GTGAGTACTC AACCAAGTCA TTCTGAGAAT AGTGTATGCG GCGACCGAGT TGCTCTTGCC18550 18560 18570 18580 18590 18600
TGGCGTCAAC ACGGGATAAT ACCGCGCCAC ATAGCAGAAC TTTAAAAGTG CTCATCATTG18610 18620 18630 18640 18650 18660
GAAACGTTT TCGGGGCGA AAACCTCTCA GGATCTTACC GCTGTTGAGA TCCAGTTCTA18670 18680 18690 18700 18710 18720
TGTAACCCAC TCGTGCACCC AACTGATCTT CAGCATCTTT TACTTTCACC AGCGTTTCTG18730 18740 18750 18760 18770 18780
GGTGAGCAAA AACAGGAAGG CAAAATGCCG CAAAAAAGGG AATAAGGGCG ACACGGAAAT18790 18800 18810 18820 18830 18840
GTTGAATACT CATACTCTT CTTTTTCAAT ATTATTGAAG CATTATATCAG GGTTATTGTC18850 18860 18870 18880 18890 18900
TCATGAGCGG ATACATATTT GAATGTATTT AGAAAAATAA ACAAATAGGG GTTCCGCGCA18910 18920 18930 18940 18950 18960
TATTTCCCG AAAAGTGCCA CCTGACGTCT AAGAAACCAT TATTATCATG ACATTAACTT18970 18980 18990 19000 19010 19020
ATAAAAATAG GCGTATCAG AGGCCCTTTC GTCTTCAAGA A.....

Mandy + 5E8N-SHL

FIGURE 9



- Nt D = Inactive Dihydrofolate reductase
 E = CMV and SV40 enhancers
 Nt H = Inactive Salmonella Histidinol Dehydrogenase
 T = Herpes Simplex thymidine kinase promoter and polyoma enhancer
 C = Cytomegalovirus promoter/enhancer
 N1 = Neomycin phosphotransferase exon 1
 K = Human kappa constant
 VL = Variable light chain anti-CD23 primate 5E8 and leader
 VH = Variable heavy chain anti-CD23 primate 5E8N- and leader
 SO = SV40 Origin of replication
 B = Bovine growth hormone polyadenylation
 M2 = Neomycin phosphotransferase exon 2
 G1 = Human Gamma 1 constant

Mandy cut Xba I Xho I and ligated to Xba I Xho I fragment from XKG1+CD23 5E8N-SHL

Map by Mitchell Reff Constructed by Karen McLachlan 05/26/97 19,035 bp
 Noncutters = AflII, AvrII, HindIII, I-PpoI, I-SceI, PmlI, RsrII, SgfI, SrfI

FIGURE 10

DNASIS

Mandy + SE8N-SHL

```

      10      20      30      40      50      60
TTAATTAAGG GCGCGAGAAT GGGCGGAACT GGGCGGAGTT AGGGGCGGGA TGGGCGGAGT

      70      80      90     100     110     120
TAGGGGCGGG ACTATGGTTG CTGACTAATT GAGATGCATG CTTTGCATAC TTCTGCCTGC

      130     140     150     160     170     180
TGGGGAGCCT GGGGACTTTC CACACCTGGT TGCTGACTAA TTGAGATGCA TGCTTTGCAT

      190     200     210     220     230     240
ACTTCTGCCT GCTGGGGAGC CTGGGGACTT TCCACACCTT AACTGACACA CATTCCACAG

      250     260     270     280     290     300
AATTAATTCC CCTAGTTATT AATAGTAATC AATTACGGGG TCATTAGTTC ATAGCCCAT

      310     320     330     340     350     360
TATGGAGTTC CGCGTTACAT AACTTACGGT AAATGGCCCG CCTGGCTGAC CGCCCAACGA

      370     380     390     400     410     420
:CCCCCCA TTGACGTCAA TAATGACGTA TGTTCCTATA GTAACGCCAA TAGGGACTTT

      430     440     450     460     470     480
CCATTGACGT CAATGGGTGG AGTATTTACG GTAAACTGCC CACTTGGCAG TACATCAAGT

      490     500     510     520     530     540
GTATCATATG CCAAGTACGC CCCCTATTGA CGTCAATGAC GGTAAATGGC CCGCCTGGCA

      550     560     570     580     590     600
TTATGCCCAG TACATGACCT TATGGGACTT TCCTACTTGG CAGTACATCT ACGTATTAGT

      610     620     630     640     650     660
CATCGCTATT ACCATGGTGA TGGCGTTTTG GCAGTACATC AATGGGCGTG GATAGCGGTT

      670     680     690     700     710     720
TGACTCACGG GGATTTCCAA GTCTCCACCC CATTGACGTC AATGGGAGTT TGTTTTGAAG

      730     740     750     760     770     780
:GTTTAAAC AGCTTGCCCG GCCAGCTTTA TTTAACGTGT TTACGTCGAG TCAATTGTAC

      790     800     810     820     830     840
ACTAACGACA GTGATGAAAG AAATACAAAA GCGCATAATA TTTTGAACGA CGTCGAACCT

      850     860     870     880     890     900
TTATTACAAA ACAAAACACA AACGAATATC GACAAAGCTA GATTGCTGCT ACAAGATTG

      910     920     930     940     950     960
GCAAGTTTTG TGGCGTTGAG CGAAAATCCA TTAGATAGTC CAGCCATCGG TTCGGAAAAA

      970     980     990    1000    1010    1020
CAACCCCTGT TTGAAACTAA TCGAAACCTA TTTTACAAAT CTATTGAGGA TTTAATATTT

      1030    1040    1050    1060    1070    1080
AAATTCAGAT ATAAAGACGC TGAAAATCAT TTGATTTTCG CTCTAACATA CCACCCTAAA

      1090    1100    1110    1120    1130    1140
GATTATAAAT TTAATGAATT ATTAAAATAC ATCAGCAACT ATATATTGAT AGACATTTCC

      1150    1160    1170    1180    1190    1200
AGTTTGTGAT ATTAGTTTGT GCGTCTCATT ACAATGGCTG TTATTTTAA CAACAAACAA

      1210    1220    1230    1240    1250    1260
CTGCTCGCAG ACAATAGTAT AGAAAAGGGA GGTGAACTGT TTTTGTTTAA CGGTTCGTAC

      1270    1280    1290    1300    1310    1320
AACATTTTGG AAAGTTATGT TAATCCGGTG CTGCTAAAAA ATGGTGTAAT TGAAGTAGAA

```

DNASIS

Mandy + SE8N-SHL

```

      1330      1340      1350      1360      1370      1380
GAAGCTGCGT ACTATGCCGG CAACATATTG TACAAAACCG ACGATCCCAA ATTCATTGAT

      1390      1400      1410      1420      1430      1440
TATATAAATT TAATAATTAA AGCAACACAC TCCGAAGAAC TACCAGAAAA TAGCACTGTT

      1450      1460      1470      1480      1490      1500
GTAAATTACA GAAAAACTAT GCGCAGCGGT ACTATACACC CCATTAAAAA AGACATATAT

      1510      1520      1530      1540      1550      1560
ATTTATGACA AAAAAAATT TACTCTATAC GATAGATACA TATATGGATA CGATAATAAC

      1570      1580      1590      1600      1610      1620
TATGTTAATT TTTATGAGGA GAAAAATGAA AAAGAGAAGG AATACGAAGA AGAAGACGAC

      1630      1640      1650      1660      1670      1680
AAGGCGTCTA GTTTATGTGA AAATAAAATT ATATTGTCGC AAATTAAGTG TGAATCATTT

      1690      1700      1710      1720      1730      1740
GAAAAATGATT TTAATATTA CCTCAGCGAT TATAACTACG CGTTTTCAAT TATAGATAAT

      1750      1760      1770      1780      1790      1800
ACTACAAATG TTCTTGTTGC GTTTGGTTTG TATCGTTAAT AAAAAACAAA TTTGACATTT

      1810      1820      1830      1840      1850      1860
ATAATGTTT TATTATTCAA TAATTACAAA TAGGATTGAG ACCCTTGCAAG TTGCCAGCAA

      1870      1880      1890      1900      1910      1920
ACGGACAGAG CTTGTCGAGG AGAGTTGTTG ATTCATTGTT TGCTCCCTG CTGCGGTTTT

      1930      1940      1950      1960      1970      1980
TCACCGAAGT TCATGCCAGT CCAGCGTTTT TGCAGCAGAA AAGCCGCCGA CTTGCGTTTG

      1990      2000      2010      2020      2030      2040
CGGTGCGGAG TGAAGATCCC TTTCTTGTTA CCGCCAACGC GCAATATGCC TTGCGAGGTC

      2050      2060      2070      2080      2090      2100
GCAAAATCGG CGAAATTCCA TACCTGTTCA CCGACGACGG CGCTGACGCG ATCAAAGACG

      2110      2120      2130      2140      2150      2160
CGGTGATACA TATCCAGCCA TGCACACTGA TACTCTTCAC TCCACATGTC GGTGTACATT

      2170      2180      2190      2200      2210      2220
GAGTGACGCC CGGCTAACGT ATCCACGCCG TATTCGGTGA TGATAATCGG CTGATGCAGT

      2230      2240      2250      2260      2270      2280
TTCTCCTGCC AGGCCAGAAG TTCTTTTCC AGTACCTTCT CTGCCGTTTC CAAATCGCCG

      2290      2300      2310      2320      2330      2340
CTTTGGACAT ACCATCCGTA ATAACGGTTC AGGCACAGCA CATCAAAGAG ATCGCTGATG

      2350      2360      2370      2380      2390      2400
GTATCGGTGT GAGCGTCGCA GAACATTACA TTGACGCAAG TGATCGGACG CGTCGGGTCTG

      2410      2420      2430      2440      2450      2460
AGTTTACGCG TTGCTTCCGC CAGTGGCGCG AAATATTCCC GTGCACCTTG CGGACGGGTA

      2470      2480      2490      2500      2510      2520
TCCGGTTCGT TGGCAATACT CCACATCACC ACGCTTGGGT GGTTTTGTG ACGCGCTATC

      2530      2540      2550      2560      2570      2580
AGCTCTTTAA TCGCCTGTAA GTGCGCTTGC TGAGTTTCCC CGTTGACTGC CTCTTCGCTG

      2590      2600      2610      2620      2630      2640

```

DNASIS

Mandy + SE8N-SHL

TACAGTTCTT TCGGCTTGTT GCCCGCTTCG AAACCAATGC CTAAAGAGAG GTTAAAGCCG
 2650 2660 2670 2680 2690 2700
 ACAGCAGCAG TTTCATCAAT CACCACGATG CCATGTTTAT CTGCCCAGTC GAGCATCTCT
 2710 2720 2730 2740 2750 2760
 TCAGCGTAAG GGTAATGCGA GGTACGGTAG GAGTTGGCCC CAATCCAGTC CATTAATGCG
 2770 2780 2790 2800 2810 2820
 TGGTCGTGCA CCATCAGCAC GTTATCGAAT CCTTTGCCAC GCAAGTCCGC ATCTTCATGA
 2830 2840 2850 2860 2870 2880
 CGACCAAAGC CAGTAAAGTA GAACGGTTTG TGGTTAATCA GGAAGTGTTC GCCCTTCACT
 2890 2900 2910 2920 2930 2940
 GCCACTGACC GGATGCCGAC GCGAAGCGGG TAGATATCAC ACTCTGTCTG GCTTTTGGCT
 2950 2960 2970 2980 2990 3000
 TGACGCACA GTTCATAGAG ATAACCTTCA CCCGGTTGCC AGAGGTGCGG ATTCACCACT
 3010 3020 3030 3040 3050 3060
 TGCAAAGTCC CGCTAGTGCC TTGTCCAGTT GCAACCACCT GTTGATCCGC ATCACGCAGT
 3070 3080 3090 3100 3110 3120
 TCAACGCTGA CATCACCATT GGCCACCACC TGCCAGTCAA CAGACGCGTG GTTACAGTCT
 3130 3140 3150 3160 3170 3180
 TCGCGCATAT GCGTCACCAC GGTGATATCG TCCACCCAGG TGTTCCGCGT GGTGTAGAGC
 3190 3200 3210 3220 3230 3240
 ATTACGCTGC GATGGATTCC GGCATAGTTA AAGAAATCAT GGAAGTAAGA CTGCTTTTTC
 3250 3260 3270 3280 3290 3300
 TTGCCGTTTT CGTCGGTAAT CACCATTTCC GCGGGGATAG TCTGCCAGTT CAGTTCGTTG
 3310 3320 3330 3340 3350 3360
 TCACACAAA CCGTGATACC CCTCGACGGA TTAAAGACTT CAAGCGGTCA ACTATGAAGA
 3370 3380 3390 3400 3410 3420
 AGTGTTCTGC TTCGTCCCAG TAAGCTATGT CTCCAGAATG TAGCCATCCA TCCTTGTCAG
 3430 3440 3450 3460 3470 3480
 TCAAGGCGTT GGTCGCTTCC GGATTGTTTA CATAACCGGA CATAATCATA GGTCTCTCTGA
 3490 3500 3510 3520 3530 3540
 CACATAATTC GCCTCTCTGA TTAACGCCCA GCGTTTTCCC GGTATCCAGA TCCACAACCT
 3550 3560 3570 3580 3590 3600
 TCGCTTCAAA AAATGGAACA ACTTTACCGA CCGCGCCCGG TTTATCATCC CCCTCGGGTG
 3610 3620 3630 3640 3650 3660
 TAATCAGAAT AGCTGATGTA GTCTCAGTGA GCCCATATCC TTGTCGTATC CCTGGAAGAT
 3670 3680 3690 3700 3710 3720
 GGAAGCGTTT TGCAACCGCT TCCCCGACTT CTTTCGAAAG AGGTGCGCCC CCAGAAGCAA
 3730 3740 3750 3760 3770 3780
 TTTCGTGTAA ATTAGATAAA TCGTATTTGT CAATCAGAGT GCTTTTGGCG AAGAATGAAA
 3790 3800 3810 3820 3830 3840
 ATAGGGTTGG TACTAGCAAC GCACTTTGAA TTTTGTAAAT CTGAAGGGAT CGTAAAAACA
 3850 3860 3870 3880 3890 3900
 GCTCTTCTTC AAATCTATAC ATTAAGACGA CTCGAAATCC ACATATCAAA TATCCGAGTG

DNASIS

Mandy + SE8N-SHL

```

      3910      3920      3930      3940      3950      3960
TAGTAAACAT TCCAAAACCG TGATGGAATG GAACAACACT TAAATCGCA GTATCCGGAA

      3970      3980      3990      4000      4010      4020
TGATTGATT GCCAAAATA GGATCTCTGG CATGCGAGAA TCTGACGAG GCAGTTCTAT

      4030      4040      4050      4060      4070      4080
GCGGAAGGGC CACACCCTTA GGTAACCCAG TAGATCCAGA GGAATTGTTT TGTACGATC

      4090      4100      4110      4120      4130      4140
AAAGGACTCT GGTACAAAAT CGTATTCATT AAAACCGGGA GGTAGATGAG ATGTGACGAA

      4150      4160      4170      4180      4190      4200
CGTGTACATC GACTGAAATC CCTGGTAATC CGTTTTAGAA TCCATGATAA TAATTTTCTG

      4210      4220      4230      4240      4250      4260
GATTATTGGT AATTTTTTTT GCACGTTCAA AATTTTTTGC AACCCCTTTT TGGAAACAAA

      4270      4280      4290      4300      4310      4320
CTACGGTA GGCTGCGAAA TGTCATACT GTTGAGCAAT TCACGTCAT TATAAATGTC

      4330      4340      4350      4360      4370      4380
GTTGCGGGGC GCAACTGCAA CTCGATAAA TAACGCGCCC AACACCGGCA TAAAGAATTG

      4390      4400      4410      4420      4430      4440
AAGAGAGTTT TCACTGCATA CGACGATTCT GTGATTGTGA TTCAGCCCAT ATCGTTTCAT

      4450      4460      4470      4480      4490      4500
AGCTTCTGCC AACCGAACGG ACATTTCGAA GTATTCCGCG TACAGCCCGG CCGTTTAAAC

      4510      4520      4530      4540      4550      4560
GGCCGGGCTT CAATACCCTG ATTGACTGGA ACAGCTGTAG CCCTGAACAG CAGCGTGCGC

      4570      4580      4590      4600      4610      4620
TGCTGACGCG TCCGGCGATT TCCGCTCTG ACAGTATTAC CCGGACGGTC AGCGATATTC

      4630      4640      4650      4660      4670      4680
GATAATGT AAAAAACGCG GGTGACGATG CCCTGCGTGA ATACAGCGCT AAATTTGATA

      4690      4700      4710      4720      4730      4740
AAACAGAAAGT GACAGCGCTA CGCGTCACCC CTGAAGAGAT CGCCGCCGCC GCGCGCGCTC

      4750      4760      4770      4780      4790      4800
TGAGCGACGA ATTAACACAG GCGATGACCG CTGCCGTCAA AAATATTGAA ACGTTCCATT

      4810      4820      4830      4840      4850      4860
CCGCGCAGAC GCTACCGCCT GTAGATGTGG AAACCCAGCC AGGCGTGCGT TGCCAGCAGG

      4870      4880      4890      4900      4910      4920
TTACGCGTCC CGTCTCGTCT GTCGGTCTGT ATATTCCCGG CGGCTCGGCT CCGCTCTTCT

      4930      4940      4950      4960      4970      4980
CAACGGTGCT GATGCTGGCG ACGCCGGCGC GCATTGCGGG ATGCCAGAAG GTGGTTCTGT

      4990      5000      5010      5020      5030      5040
GCTCGCGGCC GCCCATCGCT GATGAAATCC TCTATGCGGC GCAACTGTGT GGCCTGCAGG

      5050      5060      5070      5080      5090      5100
AAATCTTTAA CGTCGGCGGC GCGCAGGCGA TTGCCGCTCT GGCCTTCGGC AGCGAGTCCG

      5110      5120      5130      5140      5150      5160
TACCGAAAGT GGATAAAATT TTTGGCCCCG GCAACGCCTT TGTAACCGAA GCCAAACGTC

      5170      5180      5190      5200      5210      5220
AGGTCAGCCA GCGTCTCGAC GGC GCGGCTA TCGATATGCC AGCCGGGCGG TCTGAAGTAC

```

DNASIS

Mandy + SE8N-SHL

```

      5230      5240      5250      5260      5270      5280
TGGTGATCGC AGACAGCGGC GCAACACCGG ATTTCGTCGC TTCTGACCTG CTCTCCCAGG

      5290      5300      5310      5320      5330      5340
CTGAGCACGG CCCGGATTCC CAGGTGATCC TGCTGACGCC TGATGCTGAC ATTGCCCGCA

      5350      5360      5370      5380      5390      5400
AGGTGGCGGA GCGCGTAGAA CGTCAACTGG CGGAACTGCC GCGCGCGGAC ACCGCCCGGC

      5410      5420      5430      5440      5450      5460
AGGCCCTGAG CGCCAGTCGT CTGATTGTGA CCAAAGATTT AGCGCAGTGC GTCGCCATCT

      5470      5480      5490      5500      5510      5520
CTAATCAGTA TGGGCCGGAA CACTTAATCA TCCAGACGCG CAATGCGCGC GATTTGGTGG

      5530      5540      5550      5560      5570      5580
ATGCGATTAC CAGCGCAGGC TCGGTATTTT TCGGCGACTG GTCGCCGGAA TCCGCCGGTG

      5590      5600      5610      5620      5630      5640
ATTACGCTTC CGGAACCAAC CATGTTTTAC CGACCTATGG CTATACTGCT ACCTGTTCCA

      5650      5660      5670      5680      5690      5700
GCCTTGGGTT AGCGGATTTT CAGAAACGGA TGACCGTTCA GGAAGTGTCT AAAGCGGGCT

      5710      5720      5730      5740      5750      5760
TTTCCGCTCT GGCATCAACC ATTGAAACAT TGGCGGGCGC AGAACGTCTG ACCGCCCATG

      5770      5780      5790      5800      5810      5820
AAAATGCCGT GACCCTGCGC GTAAACGCCC TCAAGGAGCA AGCATGAGCA CTGAAACAC

      5830      5840      5850      5860      5870      5880
TCTCAGCGTC GCTGACTTAG CCCGTGAAAA TGTCGCAAC CTGGAGATCC AGACATGGAT

      5890      5900      5910      5920      5930      5940
AAGATACATT GATGAGTTTG GACAAACCAC AACTAGAATG CAGTGAAAAA AATGCTTTAT

      5950      5960      5970      5980      5990      6000
TTGTGAAATT TGTGATGCTA TTGCTTTATT TGTAACCATT ATAAGCTGCA ATAAACAAGT

      6010      6020      6030      6040      6050      6060
TAACAACAAC AATTGCATTG ATTTTATGTT TCAGGTTTCA GGGGAGGTGT GGGAGGTTTT

      6070      6080      6090      6100      6110      6120
TTAAAGCAAG TAAACCTCT ACAATGTGG TATGGCTGAT TATGATCTCT AGGGCCGGCC

      6130      6140      6150      6160      6170      6180
CTCGACGGCG CGTCTAGAGC AGTGTGGTTT TCAAGAGGAA GCAAAAAGCC TCTCCACCCA

      6190      6200      6210      6220      6230      6240
GGCCTGGAAT GTTCCACCC AATGTCGAGC AGTGTGGTTT TGCAAGAGGA AGCAAAAAGC

      6250      6260      6270      6280      6290      6300
CTCTCACCCC AGGCCTGGAA TGTTCACCC CAATGTCGAG CAAACCCCGC CCAACGCTTT

      6310      6320      6330      6340      6350      6360
GTCATTGGCG AATTGGAACA CGCATATGCA GTCGGGGCGG CGCGGTCCCA GGTCCACTTC

      6370      6380      6390      6400      6410      6420
GCATATTAAG GTGGCGCGTG TGGCTCGAA CACCGAGCGA CCCTGCAGCC AATATGGGAT

      6430      6440      6450      6460      6470      6480
CGGCCATTGA ACAAGATGGA TTGCACGCAG GTTCTCCGGC CGCTTGGGTG GAGAGGCTAT

      6490      6500      6510      6520      6530      6540

```

DNASIS

Mandy + SE8N-SHL

TCGGCTATGA CTGGGCACAA CAGACAATCG GCTGCTCTGA TGCCGCCGTG TTCCGGCTGT

6550 6560 6570 6580 6590 6600
CAGCGCAGGG GCGCCCGGTT CTTTTGTCA AGACCGACCT GTCCGGTGCC CTGAATGAAC

6610 6620 6630 6640 6650 6660
TGCAGGTAAG TCGGCCCGTC GATGGCCGAG GCGGCCTCGG CCTCTGCATA AATAAAAAAA

6670 6680 6690 6700 6710 6720
ATTAGTCAGC CATGCATGGG GCGGAGAAATG GCGGGAATG GCGCGAGTTA GGGGCGGGAT

6730 6740 6750 6760 6770 6780
GGGCGGAGTT AGGGGCGGGA CTATGGTTGC TGAATAATTG AGATGCATGC TTTGCATACT

6790 6800 6810 6820 6830 6840
TCTGCCTGCT GGGGAGCCTG GGGACTTTCC ACACCTGGTT GCTGACTAAT TGAGATGCAT

6850 6860 6870 6880 6890 6900
CCTTGCATA CTTCTGCCTG CTGGGGAGCC TGGGGACTTT CCACACCCTA ACTGACACAC

6910 6920 6930 6940 6950 6960
ATTCCACAGA ATTAATTCCC CTAGTTATTA ATAGTAATCA ATTACGGGGT CATTAGTTCA

6970 6980 6990 7000 7010 7020
TAGCCCATAT ATGGAGTTCC GCGTTACATA ACTTACGGTA AATGGCCCGC CTGGCTGACC

7030 7040 7050 7060 7070 7080
GCCCAACGAC CCCC GCCCAT TGACGTCAAT AATGACGTAT GTTCCCATAG TAACGCCAAT

7090 7100 7110 7120 7130 7140
AGGGACTTTC CATTGACGTC AATGGGTGGA GTATTACGG TAACTGCCC ACTTGGCAGT

7150 7160 7170 7180 7190 7200
ACATCAAGTG TATCATATGC CAAGTACGCC CCCTATTGAC GTCAATGACG GTAAATGGCC

7210 7220 7230 7240 7250 7260
CGCTGGCAT TATGCCCAGT ACATGACCTT ATGGGACTTT CTTACTTGGC AGTACATCTA

7270 7280 7290 7300 7310 7320
CGTATTAGTC ATCGCTATTA CCATGGTGAT GCGGTTTTGG CAGTACATCA ATGGGCGTGG

7330 7340 7350 7360 7370 7380
ATAGCGGTTT GACTCACGGG GATTTCCAAG TCTCCACCCC ATTGACGTCA ATGGGAGTTT

7390 7400 7410 7420 7430 7440
GTTTTGGCAC CAAAATCAAC GGGACTTTCC AAAATGTCGT AACAACTCCG CCCCATTGAC

7450 7460 7470 7480 7490 7500
GCAAATGGGC GGTAGGCGTG TACGGTGGGA GGTCTATATA AGCAGAGCTG GGTACGTGAA

7510 7520 7530 7540 7550 7560
CCGTCAGATC GCCTGGAGAC GCCATCACAG ATCTCTCACC ATGGACATGA GGGTCCCCGC

7570 7580 7590 7600 7610 7620
TCAGCTCCTG GGGCTCCTTC TGCTCTGGCT CCCAGGTGCC AGATGTGACA TCCAGATGAC

7630 7640 7650 7660 7670 7680
CCAGTCTCCA TCTTCCCTGT CTGCATCTGT AGGGGACAGA GTCACCATCA CTTGCAGGGC

7690 7700 7710 7720 7730 7740
AAGTCAGGAC ATTAGGTATT ATTTAAATTG GTATCAGCAG AAACCAGGAA AAGCTCCTAA

7750 7760 7770 7780 7790 7800
GCTCCTGATC TATGTTGCAT CCAGTTTGCA AAGTGGGGTC CCATCAAGGT TCAGCGGCAG

DNASIS

Mandy + SE8N-SHL

```

      7810      7820      7830      7840      7850      7860
TGGATCTGGG ACAGAGTTCA CTCTACCGT CAGCAGCTG CAGCCTGAAG ATTTJGCGAC

      7870      7880      7890      7900      7910      7920
TTATTACTGT CTACAGGTTT ATAGTACCCC TCGGACGTTT GGCCAAGGGA CCAAGGTGGA

      7930      7940      7950      7960      7970      7980
AATCAAACGT ACGGTGGCTG CACCATCTGT CTTTCATCTT CCGCCATCTG ATGAGCAGTT

      7990      8000      8010      8020      8030      8040
GAAATCTGGA ACTGCCTCTG TTGTGTGCCT GCTGAATAAC TTCTATCCCA GAGAGGCCAA

      8050      8060      8070      8080      8090      8100
AGTACAGTGG AAGGTGGATA ACGCCCTCCA ATCGGGTAAC TCCCAGGAGA GTGTCACAGA

      8110      8120      8130      8140      8150      8160
GCAGGACAGC AAGGACAGCA CCTACAGCCT CAGCAGCACC CTGACGCTGA GCAAAGCAGA

      8170      8180      8190      8200      8210      8220
ACGAGAAA CACAAAGTCT ACGCCTGCGA AGTCACCCAT CAGGGCCTGA GCTCGCCCGT

      8230      8240      8250      8260      8270      8280
CACAAAGAGC TTCAACAGGG GAGAGTGTG AATTCAGATC CGTTAACGGT TACCAACTAC

      8290      8300      8310      8320      8330      8340
CTAGACTGGA TTCGTGACAA CATGCGGCCG TGATATCTAC GTATGATCAG CCTCGACTGT

      8350      8360      8370      8380      8390      8400
GCCTTCTAGT TGCCAGCCAT CTGTTGTTG CCCCTCCCC GTGCCCTTCT TGACCTGGA

      8410      8420      8430      8440      8450      8460
AGGTGCCACT CCCACTGTCC TTCTCTAATA AAATGAGGAA ATTGCATCGC ATTGTCTGAG

      8470      8480      8490      8500      8510      8520
TAGGTGTCAT TCTATTCTGG GGGGTGGGGT GGGGCAGGAC AGCAAGGGGG AGGATTGGGA

      8530      8540      8550      8560      8570      8580
ACAATAGC AGGCATGCTG GGGATGCGGT GGGCTCTATG GCTTCTGAGG CGGAAAGAAC

      8590      8600      8610      8620      8630      8640
CAGCTGGGAC TAGTCGCAAT TGGGCGGAGT TAGGGGCGGG ATGGGCGGAG TTAGGGGCGG

      8650      8660      8670      8680      8690      8700
GACTATGGTT GCTGACTAAT TGAGATGCAT GCTTTGCATA CTTCTGCTG CTGGGGAGCC

      8710      8720      8730      8740      8750      8760
TGGGGACTTT CCACACCTGG TTGCTGACTA ATTGACATGC ATGCTTTGCA TACTTCTGCC

      8770      8780      8790      8800      8810      8820
TGCTGGGGAG CCTGGGGACT TTCCACACCC TAACTGACAC ACATTCCACA GAATTAATTC

      8830      8840      8850      8860      8870      8880
CCCTAGTTAT TAATAGTAAT CAATTACGGG GTCATTAGTT CATAGCCCAT ATATGGAGTT

      8890      8900      8910      8920      8930      8940
CCGCGTTACA TAACTTACGG TAAATGGCCC GCCTGGCTGA CCGCCCAACG ACCCCCGCCC

      8950      8960      8970      8980      8990      9000
ATTGACGTCA ATAATGACGT ATGTTCCCAT AGTAACGCCA ATAGGGACTT TCCATTGACG

      9010      9020      9030      9040      9050      9060
TCAATGGGTG GAGTATTTAC GGTAACTGC CCACTTGGCA GTACATCAAG TGTATCATAT

      9070      9080      9090      9100      9110      9120
GCCAAGTACG CCCCCTATTG ACGTCAATGA CGGTAAATGG CCCGCTGGC ATTATGCCCA

```

DNASIS
Mandy + SE8N-SHL

```

      9130      9140      9150      9160      9170      9180
GTACATGACC TTATGGGACT TTCCTACTTG GCAGTACATC TACGTATTAG TCATCGCTGT

      9190      9200      9210      9220      9230      9240
TACCATGGTG ATGCGGTTTT GGCAGTACAT CAATGGGCGT GGATAGCGGT TTGACTCAGC

      9250      9260      9270      9280      9290      9300
GGGATTTCCA AGTCTCCACC CCATTGACGT CAATGGGAGT TTGTTTTGGC ACCAAAATCA

      9310      9320      9330      9340      9350      9360
ACGGGACTTT CCAAAATGTC GTAACAACTC CGCCCCATTG ACGCAAATGG GCGGTAGGCG

      9370      9380      9390      9400      9410      9420
TGTACGGTGG GAGGTCTATA TAAGCAGAGC TGGGTACGTG AACCGTCAGA TCGCCTGGAG

      9430      9440      9450      9460      9470      9480
ACGCCGTCGA CATGGGTTGG AGCCTCATCT TGCTCTTCCT TGTCGCTGTT GCTACGCGTG

      9490      9500      9510      9520      9530      9540
.CCTGTCCGA GGTGCAGCTG GTGGAGTCTG GGGGCGGCTT GGCAAAGCCT GGGGGGTCCC

      9550      9560      9570      9580      9590      9600
TGAGACTCTC CTGCGCAGCC TCCGGGTTCA GGTTCACTT CAATAACTAC TACATGGACT

      9610      9620      9630      9640      9650      9660
GGGTCCGCCA GGCTCCAGGG CAGGGGCTGG AGTGGGTCTC ACGTATTAGT AGTAGTGGTG

      9670      9680      9690      9700      9710      9720
ATCCCACATG GTACGCAGAC TCCGTGAAGG GCAGATTCAC CATCTCCAGA GAGAACGCCA

      9730      9740      9750      9760      9770      9780
AGAACACACT GTTCTTCAA ATGAACAGCC TGAGAGCTGA GGACACGGCT GTCTATTACT

      9790      9800      9810      9820      9830      9840
GTGCGAGCTT GACTACAGGG TCTGACTCCT GGGGCCAGGG AGTCTGCTC ACCGTCTCCT

      9850      9860      9870      9880      9890      9900
LAGCTAGCAC CAAGGGCCCA TCGGTCTTCC CCCTGGCACC CTCCTCCAAG AGCACCTCTG

      9910      9920      9930      9940      9950      9960
GGGGCAGAGC GGCCCTGGGC TGCCTGGTCA AGGACTACTT CCCCGAACCG GTGACGGTGT

      9970      9980      9990      10000      10010      10020
CGTGGAACTC AGGCGCCCTG ACCAGCGGCG TGCAACCTT CCCGGCTGTC CTACAGTCCT

      10030      10040      10050      10060      10070      10080
CAGGACTCTA CTCCTCAGC AGCGTGGTGA CCGTGCCCTC CAGCAGCTTG GGCACCCAGA

      10090      10100      10110      10120      10130      10140
CCTACATCTG CAACGTGAAT CACAAGCCCA GCAACACCAA GGTGGACAAG AAAGTTGAGC

      10150      10160      10170      10180      10190      10200
CCAAATCTTG TGACAAAAC CACACATGCC CACCGTGCCC AGCACCTGAA CTCCTGGGGG

      10210      10220      10230      10240      10250      10260
GACCGTCAGT CTTCTCTTC CCCCCAAAAC CCAAGGACAC CCTCATGATC TCCCGGACCC

      10270      10280      10290      10300      10310      10320
CTGAGGTCAC ATGCGTGGTG GTGGACGTGA GCCACGAAGA CCCTGAGGTC AAGTTCAACT

      10330      10340      10350      10360      10370      10380
GGTACGTGGA CGGCGTGGAG GTGCATAATG CCAAGACAAA GCCGCGGGAG GAGCAGTACA

      10390      10400      10410      10420      10430      10440

```

DNASIS

Mandy + SE8N-SHL

ACAGCACGTA CCGTGTGGTC AGCGTCCTCA CCGTCCTGCA CCAGGACTGG CTGAATGGCA
 10450 10460 10470 10480 10490 10500
 AGGAGTACAA GTGCAAGGTC TCCAACAAAG CCCTCCCAGC CCCCATCGAG AAAACCATCT
 10510 10520 10530 10540 10550 10560
 CCAAAGCCAA AGGGCAGCCC CGAGAACCAC AGGTGTACAC CTTGCCCCCA TCCCGGGATG
 10570 10580 10590 10600 10610 10620
 AGCTGACCAA GAACCAGGTC AGCCTGACCT GCCTGGTCAA AGGCTTCTAT CCCAGCGACA
 10630 10640 10650 10660 10670 10680
 TCGCCGTGGA GTGGGAGAGC AATGGGCAGC CGGAGAACAA CTACAAGACC ACGCCTCCCG
 10690 10700 10710 10720 10730 10740
 TGCTGGACTC CGACGGCTCC TTCTTCTCT ACAGCAAGCT CACCGTGGAC AAGAGCAGGT
 10750 10760 10770 10780 10790 10800
 CGCAGCAGGG GAACGTCTTC TCATGCTCCG TGATGCATGA GGCTCTGCAC AACCACTACA
 10810 10820 10830 10840 10850 10860
 CGCAGAAGAG CCTCTCCCTG TCTCCGGGTA AATGAGGATC CGTTAACGGT TACCAACTAC
 10870 10880 10890 10900 10910 10920
 CTAGACTGGA TTCGTGACAA CATGCGGCCG TGATATCTAC GTATGATCAG CCTCGACTGT
 10930 10940 10950 10960 10970 10980
 GCCTTCTAGT TGCCAGCCAT CTGTTGTTTG CCCCTCCCC GTGCCTTCCT TGACCTGGGA
 10990 11000 11010 11020 11030 11040
 AGGTGCCACT CCCACTGTCC TTTCCTAATA AAATGAGGAA ATTGCATCGC ATTGTCTGAG
 11050 11060 11070 11080 11090 11100
 TAGGTGTCT TCTATTCTGG GGGGTGGGGT GGGGCAGGAC AGCAAGGGGG AGGATTGGGA
 11110 11120 11130 11140 11150 11160
 ACAATAGC AGGCATGCTG GGGATGCGGT GGGCTCTATG GCTTCTGAGG CGGAAAGAAC
 11170 11180 11190 11200 11210 11220
 CAGCTGGGGC TCGACAGCAA CGCTAGGTCTG AGGCCGCTAC TAACTCTCTC CTCCTCTCTT
 11230 11240 11250 11260 11270 11280
 TTCTCTGAG GACGAGGCAG CGCGGCTATC GTGGCTGGCC ACGACGGGCG TTCCTTGCGC
 11290 11300 11310 11320 11330 11340
 AGCTGTGCTC GACGTTGTCA CTGAAGCGGG AAGGGACTGG CTGCTATTGG GCGAAGTGCC
 11350 11360 11370 11380 11390 11400
 GGGGCAGGAT CTCCTGTCTC CTCACCTTGC TCCTGCGGAG AAAGTATCCA TCATGGCTGA
 11410 11420 11430 11440 11450 11460
 TGCAATGCGG CGGCTGCATA CGCTTGATCC GGCTACCTGC CCATTGACCC ACCAAGCGAA
 11470 11480 11490 11500 11510 11520
 ACATCGCATC GAGCGAGCAC GTACTCGGAT GGAAGCCGGT CTTGTGATC AGGATGATCT
 11530 11540 11550 11560 11570 11580
 GGACGAAGAG CATCAGGGGC TCGCGCCAGC CGAACTGTTT GCCAGGTAAG TGAGCTCCAA
 11590 11600 11610 11620 11630 11640
 TTCAAGCTCT CGAGCTAGGG CGGCCAGCTA GTAGCTTTGC TTCTCAATTT CTTATTGCA
 11650 11660 11670 11680 11690 11700
 TAATGAGAAA AAAAGGAAAA TTAATTTTAA CACCAATTCA GTAGTTGATT GAGCAAATGC

DNASIS
Mandy + SE8N-SHL

| | | | | | |
|------------|-------------|------------|-------------|------------|------------|
| 11710 | 11720 | 11730 | 11740 | 11750 | 11760 |
| GTTGCCAAAA | AGGATGCTTT | AGAGACAGTG | TTCTCTGCAC | AGATAAGGAC | AAACATTATT |
| 11770 | 11780 | 11790 | 11800 | 11810 | 11820 |
| CAGAGGGAGT | ACCCAGAGCT | GAGACTCCTA | AGCCAGTGAG | TGGCACAGCA | TCCAGGGAGA |
| 11830 | 11840 | 11850 | 11860 | 11870 | 11880 |
| AATATGCTTG | TCATCACC GA | AGCCTGATTC | CGTAGAGCCA | CACCCTGGTA | AGGGCCAATC |
| 11890 | 11900 | 11910 | 11920 | 11930 | 11940 |
| TGCTCACACA | GGATAGAGAG | GGCAGGAGCC | AGGGCAGAGC | ATATAAGGTG | AGGTAGGATC |
| 11950 | 11960 | 11970 | 11980 | 11990 | 12000 |
| AGTTGCTCCT | CACATTTGCT | TCTGACATAG | TTGTGTTGGG | AGCTTGGATA | GCTTGGGGGG |
| 12010 | 12020 | 12030 | 12040 | 12050 | 12060 |
| GGGACAGCTC | AGGGCTGCGA | TTTCGCGCCA | AACTTGACGG | CAATCCTAGC | GTGAAGGCTG |
| 12070 | 12080 | 12090 | 12100 | 12110 | 12120 |
| *AGGATTTT | ATCCCCGCTG | CCATCATGGT | TCGACCATTG | AACTGCATCG | TCGCCGTGTC |
| 12130 | 12140 | 12150 | 12160 | 12170 | 12180 |
| .CCAAATATG | GGGATTGGCA | AGAACGGAGA | CCTACCCTGG | CCTCCGCTCA | GGAACGAGTT |
| 12190 | 12200 | 12210 | 12220 | 12230 | 12240 |
| CAAGTACTTC | CAAAGAATGA | CCACAACCTC | TTCACTGGAA | GGTAAACAGA | ATCTGGTGAT |
| 12250 | 12260 | 12270 | 12280 | 12290 | 12300 |
| TATGGGTAGG | AAAACCTGGT | TCTCCATTCC | TGAGAAGAAT | CGACCTTTAA | AGGACAGAAT |
| 12310 | 12320 | 12330 | 12340 | 12350 | 12360 |
| TAATATAGTT | CTCAGTAGAG | AACTCAAAGA | ACCACCACGA | GGAGCTCATT | TTCTTGCCAA |
| 12370 | 12380 | 12390 | 12400 | 12410 | 12420 |
| AAGTTTGAT | GATGCCTTAA | CGTAGGCGCG | CCATTAAAGAC | TTATTGAACA | ACCGGAATTG |
| 12430 | 12440 | 12450 | 12460 | 12470 | 12480 |
| *AAGTAAAG | TAGACATGGT | TTGGATAGTC | GGAGGCAGTT | CTGTTTACCA | GGAAGCCATG |
| 12490 | 12500 | 12510 | 12520 | 12530 | 12540 |
| AATCAACCAG | GCCACCTCAG | ACTCTTTGTG | ACAAGGATCA | TGCAGGAATT | TGAAAGTGAC |
| 12550 | 12560 | 12570 | 12580 | 12590 | 12600 |
| ACGTTTTTCC | CAGAAATTGA | TTTGGGGAAA | TATAAACTTC | TCCAGAATA | CCCAGGCGTC |
| 12610 | 12620 | 12630 | 12640 | 12650 | 12660 |
| CTCTCTGAGG | TCCAGGAGGA | AAAAGGCATC | AAGTATAAGT | TTGAAGTCTA | CGAGAAGAAA |
| 12670 | 12680 | 12690 | 12700 | 12710 | 12720 |
| GACTAACAGG | AAGATGCTTT | CAAGTTCTCT | GCTCCCTCC | TAAAGCTATG | CATTTTTATA |
| 12730 | 12740 | 12750 | 12760 | 12770 | 12780 |
| AGACCATGGG | ACTTTTGCTG | GCTTTAGATC | AGCCTCGACT | GTGCCTTCTA | GTTGCCAGCC |
| 12790 | 12800 | 12810 | 12820 | 12830 | 12840 |
| ATCTGTTGTT | TGCCCTCCC | CCGTGCCCTC | CTTGACCTCG | GAAGGTGCCA | CTCCCACTGT |
| 12850 | 12860 | 12870 | 12880 | 12890 | 12900 |
| CCTTTCCTAA | TAAATGAGG | AAATTGCATC | GCATTGTCTG | AGTAGGTGTC | ATTCTATTCT |
| 12910 | 12920 | 12930 | 12940 | 12950 | 12960 |
| GGGGGGTGGG | GTGGGGCAGG | ACAGCAAGGG | GGAGGATTGG | GAAGACAATA | GCAGGCATGC |
| 12970 | 12980 | 12990 | 13000 | 13010 | 13020 |
| TGGGGATGCG | GTGGGCTCTA | TGGCTTCTGA | GGCGGAAAGA | ACCAGCTGGG | GCTCGAAGCG |

DNASIS
Mandy + SE8N-SHL

```

13030      13040      13050      13060      13070      13080
GCCGCCCAT TCGCTGGTGG TCAGATGCGG GATGGCGTGG GACGCGGCGG GGAGCGTCAC

13090      13100      13110      13120      13130      13140
ACTGAGGTTT TCCGCCAGAC GCCACTGCTG CCAGGCGCTG ATGTGCCCCG CTTCTGACCA

13150      13160      13170      13180      13190      13200
TGCGGTGCGG TTCGGTTGCA CTACGCGTAC TGTGAGCCAG AGTTGCCCCG CGCTCTCCGG

13210      13220      13230      13240      13250      13260
CTGCGGTAGT TCAGGCAGTT CAATCAACTG TTTACCTTGT GGAGCGACAT CCAGAGGCAC

13270      13280      13290      13300      13310      13320
TTCACCGCTT GCCAGCGGCT TACCATCCAG CGCCACCATC CAGTGCAAGG GCTCGTTATC

13330      13340      13350      13360      13370      13380
GCTATGACGG AACAGGTATT CGCTGGTCAC TTCGATGGTT TGCCCGGATA AACGGAACTG

13390      13400      13410      13420      13430      13440
AAAAACTGC TGCTGGTGTT TTGCTTCCGT CAGCGCTGGA TGCGGCGTGC GGTCGGCAAA

13450      13460      13470      13480      13490      13500
GACCAGACCG TTCATACAGA ACTGGCGATC GTTCGGCGTA TCGCCAAAAT CACCGCCGTA

13510      13520      13530      13540      13550      13560
AGCCGACCAC GGGTTGCCGT TTTCATCATA TTTAATCAGC GACTGATCCA CCCAGTCCCA

13570      13580      13590      13600      13610      13620
GACGAAGCCG CCCTGTAAAC GGGGATACTG ACGAAACGCC TGCCAGTATT TAGCGAAACC

13630      13640      13650      13660      13670      13680
GCCAAGACTG TTACCCATCG CGTGGGCGTA TTCGCAAAGG ATCAGCGGGC GCGTCTCTCC

13690      13700      13710      13720      13730      13740
AGGTAGCGAA AGCCATTTTT TGATGGACCA TTTCGGCACA GCCGGGAAGG GCTGGTCTTC

13750      13760      13770      13780      13790      13800
ATCCACGCGC GCGTACATCG GGCAAATAAT ATCGGTGGCC GTGGTGTCGG CTCCGCCGCC

13810      13820      13830      13840      13850      13860
TTCATACTGC ACCGGGCGGG AAGGATCGAC AGATTTGATC CAGCGATACA GCGCGTCGTG

13870      13880      13890      13900      13910      13920
ATTAGCGCCG TGGCCTGATT CATTCCCCAG CGACCAGATG ATCACACTCG GGTGATTACG

13930      13940      13950      13960      13970      13980
ATCGCGCTGC ACCATTGCGG TTACGCGTTC GCTCATCGCC GGTAGCCAGC GCGGATCATC

13990      14000      14010      14020      14030      14040
GGTCAGACGA TTCATTGGCA CCATGCCGTG GGTTCATAA TTGGCTTCAT CCACCACATA

14050      14060      14070      14080      14090      14100
CAGGCCGTAG CGGTGCGACA GCGTGTAACA CAGCGGATGG TTCGGATAAT GCGAACAGCG

14110      14120      14130      14140      14150      14160
CACGGCGTTA AAGTTGTTCT GCTTCATCAG CAGGATATCC TGCACCATCG TCTGCTCATC

14170      14180      14190      14200      14210      14220
CATGACCTGA CCATGCAGAG GATGATGCTC GTGACGGTTA ACGCCTCGAA TCAGCAACGG

14230      14240      14250      14260      14270      14280
CTTGCCGTTC AGCAGCAGCA GACCATTTTC AATCCGCACC TCGCGGAAAC CGACATCGCA

14290      14300      14310      14320      14330      14340

```

DNASIS

Mandy + SE8N-SHL

GGCTTCTGCT TCAATCAGCG TGCCGTCGGC GGTGTGCAGT TCAACCACCG CACGATAGAG

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 14350 | 14360 | 14370 | 14380 | 14390 | 14400 |
| ATTCGGGATT | TCGGCGCTCC | ACAGTTTCGG | GTTTTCGACG | TTCAGACGTA | GTGTGACGCG |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 14410 | 14420 | 14430 | 14440 | 14450 | 14460 |
| ATCGGCATAA | CCACCACGCT | CATCGATAAT | TTCACCGCCG | AAAGGCGCGG | TGCCGCTGGC |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 14470 | 14480 | 14490 | 14500 | 14510 | 14520 |
| GACCTGCGTT | TCACCCTGCC | ATAAAGAAAC | TGTTACCCGT | AGGTAGTCAC | GCAACTCGCC |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 14530 | 14540 | 14550 | 14560 | 14570 | 14580 |
| GCACATCTGA | ACTTCAGCCT | CCAGTACAGC | GCGGCTGAAA | TCATCATTAA | AGCGAGTGGC |

| | | | | | |
|------------|------------|------------|------------|------------|-------------|
| 14590 | 14600 | 14610 | 14620 | 14630 | 14640 |
| AACATGGAAA | TCGCTGATTT | GTGTAGTCGG | TTTATGCAGC | AACGAGACGT | CACGGAAAAAT |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 14650 | 14660 | 14670 | 14680 | 14690 | 14700 |
| CCCGCTCATC | CGCCACATAT | CCTGATCTTC | CAGATAACTG | CCGTCACTCC | AGCGCAGCAC |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 14710 | 14720 | 14730 | 14740 | 14750 | 14760 |
| CATCACCGCG | AGGCGGTTTT | CTCCGGCGCG | TAAAAATGCG | CTCAGGTCAA | ATTCAGACGG |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 14770 | 14780 | 14790 | 14800 | 14810 | 14820 |
| CAAACGACTG | TCCTGGCCGT | AACCGACCCA | GCGCCCGTTG | CACCACAGAT | GAAACGCCGA |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 14830 | 14840 | 14850 | 14860 | 14870 | 14880 |
| GTTAACGCCA | TCAAAAATAA | TTGCGCTCTG | GCCTTCTCTG | AGCCAGCTTT | CATCAACATT |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 14890 | 14900 | 14910 | 14920 | 14930 | 14940 |
| AAATGTGAGC | GAGTAACAAC | CCGTGCGATT | CTCCGTGGGA | ACAAACGGCG | GATTGACCGT |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 14950 | 14960 | 14970 | 14980 | 14990 | 15000 |
| AATGGGATAG | GTCACGTTGG | TGTAGATGGG | CGCATCGTAA | CCGTGCATCT | GCCAGTTTGA |

| | | | | | |
|-----------|------------|------------|------------|------------|------------|
| 15010 | 15020 | 15030 | 15040 | 15050 | 15060 |
| CCGACGACG | ACAGTATCGG | CCTCAGGAAG | ATCGCACTCC | AGCCAGCTTT | CCGGCACCCG |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 15070 | 15080 | 15090 | 15100 | 15110 | 15120 |
| TTCTGGTGCC | GGAAACCAGG | CAAAGCGCCA | TTCGCCATTG | AGGCTGCGCA | ACTGTTGGGA |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 15130 | 15140 | 15150 | 15160 | 15170 | 15180 |
| AGGGCGATCG | GTGCGGGCCT | CTTCGCTATT | ACGCCAGCTG | GCGAAAGGGG | GATGTGCTGC |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 15190 | 15200 | 15210 | 15220 | 15230 | 15240 |
| AAGGCGATTA | AGTTGGGTAA | CGCCAGGGTT | TTCCCACTCA | CGACGTTGTA | AAACGACTTA |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 15250 | 15260 | 15270 | 15280 | 15290 | 15300 |
| ATCCGTCGAG | GGGCTGCCTC | GAAACAGACG | ACCTTCCGTT | GTGCAGCCAG | CGGCCTCTGC |

| | | | | | |
|------------|------------|-----------|------------|------------|------------|
| 15310 | 15320 | 15330 | 15340 | 15350 | 15360 |
| GCCGGTGCCC | ACAATCGTGC | GCGAACAAC | TAAACCAGAA | CAAATTATAC | CGGCAGCACC |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 15370 | 15380 | 15390 | 15400 | 15410 | 15420 |
| GCCGCCACCA | CCTTCTCCCG | TGCTTAACAT | TCCAGCGCCT | CCACCACCAC | CACCACCATC |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 15430 | 15440 | 15450 | 15460 | 15470 | 15480 |
| GATGTCTGAA | TTGCCGCCCC | CTCCACCAAT | GCCGACGGAA | CCTCAACCCG | CTGCACCTTT |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 15490 | 15500 | 15510 | 15520 | 15530 | 15540 |
| AGACGACAGA | CAACAATTGT | TGGAAGCTAT | TAGAAACGAA | AAAAATCGCA | CTCGTCTCAG |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 15550 | 15560 | 15570 | 15580 | 15590 | 15600 |
| ACCGGTCAAA | CCAAAAACGG | CGCCCGAAAC | CAGTACAATA | GTTGAGGTGC | CGACTGTGTT |

DNASIS

Nandy + SE8N-SHL

```

15610      15620      15630      15640      15650      15660
GCCTAAAGAG ACATTGAGC CTAACCGCC GTCTGCATCA CCGCCACCAC CTCCGCTCC

15670      15680      15690      15700      15710      15720
GCCTCCGCCG CCAGCCCCGC CTGCGCTCC ACCGATGGTA GATTATCAT CAGTCCACC

15730      15740      15750      15760      15770      15780
ACCGCCGCCA TTAGTAGATT TGCGTCTGA AATGTTACCA CCGCTGCAC CATCGCTTC

15790      15800      15810      15820      15830      15840
TAACGTGTG TCTGAATTAA AATCGGGCAC AGTTAGATTG AAACCCGCC AAAAACGCC

15850      15860      15870      15880      15890      15900
GCAATCAGAA ATAATTCAA AAAGCTCAAC TACAAATTG ATCGCGGACG TGTAGCCGA

15910      15920      15930      15940      15950      15960
CACAATTAAT AGGCGTCGTG TGGCTATGGC AAAATCGTCT TCGGAAGCAA CTTCTAACGA

15970      15980      15990      16000      16010      16020
AGGGTTGG GACGACGACG ATAATCGGCC TAATAAGCT AACACGCCG ATGTTAAATA

16030      16040      16050      16060      16070      16080
TGTCGAAGCT ACTAGTGGTA CCGCTTGGCA GAACATATCC ATCGCGTCCG CCATCTCCAG

16090      16100      16110      16120      16130      16140
CAGCCGCACG CGGCGCATCT CGGGCAGCGT TGGGTCTGG CCACGGGTGC GCATGATCGT

16150      16160      16170      16180      16190      16200
GCTCCTGTG TTAGGACCC GGCTAGGCTG GCGGGGTTC CTTACTGTT AGCAGAATGA

16210      16220      16230      16240      16250      16260
ATCACCGATA CGCGAGCGAA CGTGAAGCGA CTGCTGCTGC AAAACGCTG CGACCTGAGC

16270      16280      16290      16300      16310      16320
AACAACATGA ATGTCCTCG GTTCCGTGT TTCGTAAAGT CTGGAACGC GGAAGTCAGC

16330      16340      16350      16360      16370      16380
CCTGCACC ATTATGTTCC GGATCTGCAT CGCAGGATGC TGCTGGCTAC CCTGTGGAAC

16390      16400      16410      16420      16430      16440
ACCTACATCT GTATTAACGA AGCGCTGGCA TTGACCCTGA GTGATTTTC TCTGGTCCCG

16450      16460      16470      16480      16490      16500
CCGCATCCAT ACCGCCAGTT GTTACCCTC ACAACGTTC AGTAACCGGG CATGTTTCATC

16510      16520      16530      16540      16550      16560
ATCAGTAACC CGTATCGTGA GCATCCTCTC TCGTTTCATC GGTATCATT CCCCCATGAA

16570      16580      16590      16600      16610      16620
CAGAAATCCC CTTACACGG AGGCATCAGT GACCAAACAG GAAAAACCG CCCTTAACAT

16630      16640      16650      16660      16670      16680
GCCCCGCTTT ATCAGAAGCC AGACATTAAC GCTTCTGGAG AAACCAACG AGCTGGACGC

16690      16700      16710      16720      16730      16740
GGATGAACAG GCAGACATCT GTGAATCGCT TCACGACCAC GCTGATGAGC TTTACCGCAG

16750      16760      16770      16780      16790      16800
CTGCCTCGCG CGTTTCGGTG ATGACGGTGA AAACCTCTGA CACATGCAGC TCCCGGAGAC

16810      16820      16830      16840      16850      16860
GGTCACAGCT TGTCTGTAAG CGGATGCCGG GAGCAGACAA GCGGTCAGG GCGGTCAGG

16870      16880      16890      16900      16910      16920
GGGTGTTGGC GGGTGTGGG GCGCAGCCAT GACCCAGTCA CGTAGCGATA GCGGAGTGTA

```

DNASIS

Mandy + SE8N-SHL

| | | | | | |
|-------------|------------|-------------|-------------|-------------|-------------|
| 16930 | 16940 | 16950 | 16960 | 16970 | 16980 |
| TACTGGCTTA | ACTATGCGGC | ATCAGAGCAG | ATTGTA CTGA | GAGTGCACCA | TATGCGGTGT |
| 16990 | 17000 | 17010 | 17020 | 17030 | 17040 |
| GAAATACCGC | ACAGATGCGT | AAGGAGAAAA | TACCGCATCA | GGCGCTCTTC | CGCTTCCTCG |
| 17050 | 17060 | 17070 | 17080 | 17090 | 17100 |
| CTCACTGACT | CGCTGCGCTC | GGTCGTTCCG | CTGCGCGGAG | CGGTATCAGC | TCACTCAAAG |
| 17110 | 17120 | 17130 | 17140 | 17150 | 17160 |
| GCGGTAATAC | GGTTATCCAC | AGAATCAGGG | GATAACGCAG | AAAAGAACAT | GTGAGCAAAA |
| 17170 | 17180 | 17190 | 17200 | 17210 | 17220 |
| GGCCAGCAAA | AGGCCAGGAA | CCGTAAAAAG | GCCGCGTTGC | TGGCGTTTTT | CCATAGGCTC |
| 17230 | 17240 | 17250 | 17260 | 17270 | 17280 |
| CGCCCCCTG | ACGAGCATCA | CAAAAATCGA | CGCTCAAGTC | AGAGGTGGCG | AAACCCGACA |
| 17290 | 17300 | 17310 | 17320 | 17330 | 17340 |
| GGA CTATAAA | GATACCAGGC | GTTTCCCCCT | GGAAGCTCCC | TCGTGCGCTC | TCCTGTTCCG |
| 17350 | 17360 | 17370 | 17380 | 17390 | 17400 |
| ACCCTGCCGC | TTACCGGATA | CCTGTCCGCC | TTTCTCCCTT | CGGGAAGCGT | GGCGCTTTCT |
| 17410 | 17420 | 17430 | 17440 | 17450 | 17460 |
| CATAGCTCAC | GCTGTAGGTA | TCTCAGTTCT | GTGTAGGTCG | TTGCTCCAA | GCTGGGCTGT |
| 17470 | 17480 | 17490 | 17500 | 17510 | 17520 |
| GTGCACGAAC | CCCCCGTTCA | GCCCGACCGC | TGCGCCTTAT | CCGGTA ACTA | TCGTCTTGAG |
| 17530 | 17540 | 17550 | 17560 | 17570 | 17580 |
| TCCAACCCGG | TAAGACACGA | CTTATCGCCA | CTGGCAGCAG | CCACTGCTAA | CAGGATTAGC |
| 17590 | 17600 | 17610 | 17620 | 17630 | 17640 |
| AGAGCGAGGT | ATGTAGGCGG | TGCTACAGAG | TTCTTGAAGT | GGTGGCCTAA | CTACGGCTAC |
| 17650 | 17660 | 17670 | 17680 | 17690 | 17700 |
| ACTAGAAGGA | CAGTATTTGG | TATCTGCGCT | CTGCTGAAGC | CAGTTACCTT | CGGAAAAAGA |
| 17710 | 17720 | 17730 | 17740 | 17750 | 17760 |
| GTTGGTAGCT | CTTGATCCGG | CAAA CAACC | ACCGCTGCTA | GCGGTGTTTT | TTTTGTTTGC |
| 17770 | 17780 | 17790 | 17800 | 17810 | 17820 |
| AAGCAGCAGA | TTACGCGCAG | AAAAAAAGGA | TCTCAAGAG | ATCCTTTGAT | CTTTTCTACG |
| 17830 | 17840 | 17850 | 17860 | 17870 | 17880 |
| GGGTCTGACG | CTCAGTGGAA | CGAAAACTCA | CGTTAAGGGA | TTTTGGTCAT | GAGATTATCA |
| 17890 | 17900 | 17910 | 17920 | 17930 | 17940 |
| AAAAGCATCT | TCACCTAGAT | CCTTTTAAAT | TAAAAATGAA | GTTTTAAATC | AATCTAAAGT |
| 17950 | 17960 | 17970 | 17980 | 17990 | 18000 |
| ATATATGAGT | AAACTTGCTC | TGACAGTTAC | CAATGCTTAA | TCAGTGAGGC | ACCTATCTCA |
| 18010 | 18020 | 18030 | 18040 | 18050 | 18060 |
| GCGATCTGTC | TATTTGCTTC | ATCCATAGTT | GCCTGACTCC | CCGTGCTGTA | GATAACTACG |
| 18070 | 18080 | 18090 | 18100 | 18110 | 18120 |
| ATACGGGAGG | GCTTACCATC | TGGCCCCAGT | GCTGCAATGA | TACCGCGAGA | CCCACGCTCA |
| 18130 | 18140 | 18150 | 18160 | 18170 | 18180 |
| CCGGCTCCAG | ATTTATCAGC | AATAAAC CAG | CCAGCCGGAA | GGGCCGAGCG | CAGAA GTGGT |
| 18190 | 18200 | 18210 | 18220 | 18230 | 18240 |

DNASIS

Mandy + SE8N-SHL

CCTGCAACTT TATCCGCCTC CATCCAGTCT ATTAATTGTT GCCGGGAAGC TAGAGTAACT

| | | | | | |
|-------------|------------|------------|------------|------------|------------|
| 18250 | 18260 | 18270 | 18280 | 18290 | 18300 |
| AGTTCCGCCAG | TTAATAGTTT | GCGCAACGTT | GTTGCCATTG | CTGCAGGCAT | CGTGGTGTCA |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 18310 | 18320 | 18330 | 18340 | 18350 | 18360 |
| CGCTCGTCGT | TTGGTATGGC | TTCATTGAGC | TCCGGTTCCC | AACGATCAAG | GCGAGTTACA |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 18370 | 18380 | 18390 | 18400 | 18410 | 18420 |
| TGATCCCCCA | TGTTGTGCAA | AAAAGCGGTT | AGCTCCTTCG | GTCCTCCGAT | CGTTGTGAGA |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 18430 | 18440 | 18450 | 18460 | 18470 | 18480 |
| AGTAAGTTGG | CCGCAGTGTT | ATCACTCATG | GTTATGGCAG | CACTGCATAA | TTCTCTTACT |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 18490 | 18500 | 18510 | 18520 | 18530 | 18540 |
| GTCATGCCAT | CCGTAAGATG | CTTTTCTGTG | ACTGGTGAGT | ACTCAACCAA | GTCATTCTGA |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 18550 | 18560 | 18570 | 18580 | 18590 | 18600 |
| GAATAGTGTA | TGCGGCGACC | GAGTTGCTCT | TGCCCGGCGT | CAACACGGGA | TAATACCGCG |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 18610 | 18620 | 18630 | 18640 | 18650 | 18660 |
| LCACATAGCA | GAACTTTAAA | AGTGCTCATC | ATTGGAAAAC | GTTCTTCGGG | GCGAAAACTC |

| | | | | | |
|------------|------------|------------|------------|------------|-----------|
| 18670 | 18680 | 18690 | 18700 | 18710 | 18720 |
| TCAAGGATCT | TACCGCTGTT | GAGATCCAAT | TGCATGTAAC | CCACTCGTGC | ACCCAATGA |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 18730 | 18740 | 18750 | 18760 | 18770 | 18780 |
| TCTTCAGCAT | CTTTTACTTT | CACCAGCGTT | TCTGGGTGAG | CAAAAACAGG | AAGGCAAAAT |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 18790 | 18800 | 18810 | 18820 | 18830 | 18840 |
| GCCGCAAAAA | AGGGAATAAG | GGCGACACGG | AAATGTTGAA | TACTCATACT | CTTCCTTTTT |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 18850 | 18860 | 18870 | 18880 | 18890 | 18900 |
| CAATATTATT | GAAGCATTTA | TCAGGGTTAT | TGTCTCATGA | GCGGATACAT | ATTTGAATGT |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 18910 | 18920 | 18930 | 18940 | 18950 | 18960 |
| ATTTAGAAAA | ATAAACAAAT | AGGGGTTCCG | CGCACATTTT | CCCGAAAAGT | GCCACCTGAC |

| | | | | | |
|------------|------------|------------|------------|------------|------------|
| 18970 | 18980 | 18990 | 19000 | 19010 | 19020 |
| GTCTAAGAAA | CCATTATTAT | CATGACATTA | ACCTATAAAA | ATAGGCGTAT | CACGAGGCCC |

| | | | | | |
|------------|------------|-------|-------|-------|-------|
| 19030 | 19040 | 19050 | 19060 | 19070 | 19080 |
| TTTCGTCTTC | AAGAA..... | | | | |

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 98/03935

| A. CLASSIFICATION OF SUBJECT MATTER IPC 6 C12N15/90 C12N15/85 C12Q1/68 C12N5/10 C12N9/12 C12N15/13 C07K16/28 C12N15/12 C07K14/705 G01N33/53 C12N15/62 C07K19/00 | | |
|--|--|--|
| According to International Patent Classification (IPC) or to both national classification and IPC | | |
| B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 6 C12N C12Q C07K G01N | | |
| Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched | | |
| Electronic data base consulted during the international search (name of data base and, where practical, search terms used) | | |
| C. DOCUMENTS CONSIDERED TO BE RELEVANT | | |
| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| A | WO 94 11523 A (IDEC PHARMACEUTICALS CORPORATION (US); REFF MITCHELL E. (US)) 26 May 1994 cited in the application see abstract see page 9, line 21 - page 10, line 29 see page 41, line 19 - page 42, line 19; figure 6 --- | 1,4-8, 11,12, 25-29, 31,32 |
| A | US 5 464 764 A (CAPECCHI MARIO R. AND KIRK THOMAS R.) 7 November 1995 see abstract see column 13, line 32 - column 14, line 5 --- | 1 |
| A | WO 94 05784 A (UNITED STATES AMERICA REPRESENTED BY THE SECRETARY US DPT. AGRICULTURE) 17 March 1994 see abstract --- | 1 |
| -/- | | |
| <input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex. | | |
| * Special categories of cited documents : <div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> </div> <div style="width: 45%;"> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"&" document member of the same patent family</p> </div> </div> | | |
| Date of the actual completion of the international search | | Date of mailing of the international search report |
| 23 July 1998 | | 05/08/1998 |
| Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl Fax: (+31-70) 340-3016 | | Authorized officer |
| | | Macchia, G |

INTERNATIONAL SEARCH REPORT

I. International Application No
PCT/US 98/03935

| C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|--|--|-----------------------|
| Category | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| A | WO 93 24642 A (TSI CORPORATION (US)) 9 December 1993 see abstract ---- | 1 |
| A | BARNETT R.S. ET AL.: "Antibody production in chinese hamster ovary cells using an impaired selectable marker" ACS SYMPOSIUM SERIES: ANTIBODY EXPRESSION AND ENGINEERING, vol. 604, 1995, pages 27-40, XP002072464 ----- | |

INTERNATIONAL SEARCH REPORT

I. International Application No

PCT/US 98/03935

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---------------------|----------------------------|---------------------|
| WO 9411523 A | 26-05-1994 | AU 682481 B | 09-10-1997 |
| | | AU 5613294 A | 08-06-1994 |
| | | CA 2149326 A | 26-05-1994 |
| | | DE 669986 T | 10-10-1996 |
| | | EP 0669986 A | 06-09-1995 |
| | | ES 2088838 T | 01-10-1996 |
| | | JP 8503138 T | 09-04-1996 |
| | | US 5648267 A | 15-07-1997 |
| | | US 5733779 A | 31-03-1998 |
| US 5464764 A | 07-11-1995 | US 5487992 A | 30-01-1996 |
| | | US 5627059 A | 06-05-1997 |
| | | US 5631153 A | 20-05-1997 |
| WO 9405784 A | 17-03-1994 | AU 4839493 A | 29-03-1994 |
| | | MX 9305183 A | 31-05-1994 |
| WO 9324642 A | 09-12-1993 | AU 4401993 A | 30-12-1993 |